

STIC-Biotech/ChemLib

143 104

my

From: Duffy, Patricia
Sent: Tuesday, January 25, 2005 7:26 AM
To: STIC-Biotech/ChemLib
Subject: protein sequence search 09/993,292

In re: 09/993,292

Please search SEQ ID NO:24.

Please search in commercial and interference databases.

Please print out top 50 hits.

Thanks.

Patricia A. Duffy, Ph.D.

Art Unit 1645, Remsen 3B05

571-272-0855

RECEIVED
JAN 25 2005
STIC

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: _____
Date Completed: _____
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA Sequence: # _____
AA Sequence :# _____
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

This Page Blank (uspto)

From: Chan, Christina
Sent: Tuesday, January 04, 2005 4:14 PM
To: Duffy, Patricia; STIC-Biotech/ChemLib
Subject: RE: rush sequence search

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Duffy, Patricia
Sent: Tuesday, January 04, 2005 4:08 PM
To: Chan, Christina
Subject: rush sequence search
Importance: High

Christina,

Please approve rush sequence request. Amendment due this biweek.

IN re: 09/993,292

Please search protein sequences of SEQ ID NOS:2 and 28.
Please include both an interference and commercial database search.
Please print out top 50 hits in each category.

Thank you.

Patricia A. Duffy, Ph.D.
Art Unit 1645, Remsen 3B05
571-272-0855

AA
2-305
28-303

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: _____
Date Completed: _____
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA Sequence: # _____
AA Sequence: # _____
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

Date completed: _____

Searcher: Beverly e 2528

Terminal time: _____

Elapsed time: _____

CPU time: _____

Total time: _____

Number of Searches: _____

Number of Databases: _____

Search Site

_____ STIC

_____ CM-1

_____ Pre-S

Type of Search

_____ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

✓ _____ Other CGN

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:07:40 ; Search time 39 Seconds
(without alignments)
515.240 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIFAEQTEVVKSAIETA.....TCNBYQQRHGKKTLLVDPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents_AA*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1376	91.1	309	1	US-08-557-115-3
2	1376	91.1	309	5	PCT-US94-05869-3
3	111.5	7.4	1036	4	US-09-543-681A-7736
4	109	7.2	515	4	US-09-107-532A-5317
5	108.5	7.2	1196	4	US-09-107-532A-3944
6	106.5	7.1	3696	3	US-09-134-001C-5080
7	106	7.0	1231	4	US-09-107-532A-5150
8	104	6.9	815	4	US-09-328-352-4284
9	104	6.9	1211	3	US-09-134-001C-4820
10	103	6.8	284	4	US-09-134-000C-3594
11	103	6.8	865	4	US-09-710-279-1660
12	103	6.8	961	4	US-09-914-259-66
13	102	6.8	722	4	US-09-248-796A-20613
14	101.5	6.7	1857	4	US-09-917-254-91
15	101.5	6.7	1972	4	US-09-538-092-1084
16	101	6.7	1129	4	US-09-543-681A-8019
17	100.5	6.7	526	4	US-09-447-497-15
18	100.5	6.7	553	4	US-09-447-497-14
19	100.5	6.7	566	4	US-09-538-092-807
20	100.5	6.7	609	4	US-09-447-497-12
21	100.5	6.7	613	4	US-09-447-497-9
22	100.5	6.7	623	4	US-09-447-497-11
23	100.5	6.7	627	4	US-09-447-497-8
24	100	6.6	1886	3	US-08-938-105-3
25	100	6.6	1939	3	US-09-310-187A-1
26	100	6.6	1939	4	US-09-538-092-917
27	99.5	6.6	924	4	US-09-248-796A-18798

28	99.5	6.6	1066	3	US-09-541-782-8	Sequence 8, Appli
29	99.5	6.6	1066	4	US-09-723-820-8	Sequence 8, Appli
30	99.5	6.6	1066	4	US-10-270-085-8	Sequence 8, Appli
31	99.5	6.6	2349	4	US-09-538-092-914	Sequence 914, App
32	98	6.5	618	4	US-09-134-000C-4522	Sequence 4522, Ap
33	98	6.5	1122	4	US-09-489-039A-8554	Sequence 8554, Ap
34	97.5	6.5	718	4	US-09-540-236-2753	Sequence 2753, Ap
35	97.5	6.5	804	1	US-08-785-428-2	Sequence 2, Appli
36	97.5	6.5	804	2	US-08-996-797-2	Sequence 2, Appli
37	97.5	6.5	1010	3	US-09-134-001C-5178	Sequence 5178, Ap
38	97	6.4	975	4	US-09-914-259-19	Sequence 19, Appl
39	96.5	6.4	317	4	US-09-248-796A-19212	Sequence 19212, A
40	96.5	6.4	858	4	US-09-248-796A-19055	Sequence 19055, A
41	96.5	6.4	1312	2	US-08-592-126-148	Sequence 148, App
42	96.5	6.4	1312	2	US-08-687-080-51	Sequence 51, Appl
43	96.5	6.4	1312	4	US-09-168-595-148	Sequence 148, App
44	96	6.4	360	4	US-09-248-796A-18018	Sequence 18018, A
45	96	6.4	1086	4	US-09-543-681A-7696	Sequence 7696, Ap
46	96	6.4	1960	4	US-09-538-092-1077	Sequence 1077, Ap
47	96	6.4	2871	4	US-09-538-092-936	Sequence 936, App
48	95.5	6.3	1972	4	US-08-875-435B-3	Sequence 3, Appli
49	94.5	6.3	746	3	US-08-434-000A-4	Sequence 4, Appli
50	94.5	6.3	746	3	US-09-312-157-4	Sequence 4, Appli
51	94.5	6.3	815	4	US-09-914-259-18	Sequence 18, Appl
52	94.5	6.3	1454	4	US-09-328-352-5793	Sequence 5793, Ap
53	94.5	6.3	1976	4	US-09-538-092-1078	Sequence 1078, Ap
54	94	6.2	606	4	US-08-477-831C-2	Sequence 2, Appli
55	94	6.2	631	4	US-08-477-831C-11	Sequence 11, Appl
56	94	6.2	3878	4	US-09-914-259-11	Sequence 11, Appl
57	93.5	6.2	643	4	US-09-538-092-844	Sequence 844, App
58	93.5	6.2	733	4	US-09-328-352-5599	Sequence 5599, Ap
59	93.5	6.2	829	4	US-09-248-796A-20145	Sequence 20145, A
60	93.5	6.2	1032	4	US-09-914-259-26	Sequence 26, Appl
61	93.5	6.2	1032	4	US-09-538-092-1293	Sequence 1293, Ap
62	93	6.2	451	4	US-09-134-000C-3849	Sequence 3849, Ap
63	93	6.2	885	2	US-08-533-306A-4	Sequence 4, Appli
64	93	6.2	885	2	US-08-742-923A-4	Sequence 4, Appli
65	93	6.2	10182	3	US-09-134-001C-3159	Sequence 3159, Ap
66	92.5	6.1	1027	4	US-09-914-259-27	Sequence 27, Appl
67	92.5	6.1	1111	1	US-08-317-450B-15	Sequence 15, Appl
68	92.5	6.1	1111	3	US-08-800-593-15	Sequence 15, Appl
69	92.5	6.1	1172	4	US-09-560-385A-28	Sequence 28, Appl
70	92.5	6.1	1172	4	US-09-560-385A-32	Sequence 32, Appl
71	92.5	6.1	1193	1	US-08-317-450B-13	Sequence 13, Appl
72	92.5	6.1	1193	3	US-08-800-593-13	Sequence 13, Appl
73	92.5	6.1	1193	4	US-09-560-385A-26	Sequence 26, Appl
74	92.5	6.1	1193	4	US-09-560-385A-30	Sequence 30, Appl
75	92.5	6.1	2482	1	US-08-328-254-6	Sequence 6, Appli
76	92.5	6.1	3075	2	US-08-460-309-5	Sequence 5, Appli
77	92.5	6.1	3075	2	US-08-125-077-5	Sequence 5, Appli
78	92.5	6.1	3210	4	US-09-538-092-1154	Sequence 1154, Ap
79	92.5	6.1	3248	1	US-08-353-700-1	Sequence 1, Appli
80	92.5	6.1	3248	5	PCT-US95-16216-1	Sequence 1, Appli
81	92	6.1	440	4	US-08-302-756E-35	Sequence 35, Appl
82	92	6.1	477	1	US-08-402-217A-3	Sequence 3, Appli
83	92	6.1	477	1	US-08-700-178-3	Sequence 3, Appli
84	92	6.1	477	3	US-08-995-654-3	Sequence 3, Appli
85	92	6.1	849	3	US-09-157-257-4	Sequence 4, Appli
86	92	6.1	5024	4	US-09-710-279-2964	Sequence 2964, Ap
87	91.5	6.1	350	4	US-09-489-039A-8352	Sequence 8352, Ap
88	91.5	6.1	652	4	US-09-583-110-3360	Sequence 3360, Ap
89	91.5	6.1	1939	4	US-09-538-092-915	Sequence 915, App
90	91.5	6.1	2285	3	US-09-308-375-2	Sequence 2, Appli
91	91	6.0	756	4	US-09-248-796A-17753	Sequence 17753, A
92	91	6.0	2383	4	US-09-492-709A-302	Sequence 302, App
93	90.5	6.0	451	4	US-09-538-092-706	Sequence 706, App
94	90.5	6.0	574	3	US-09-134-001C-3382	Sequence 3382, Ap
95	90.5	6.0	1937	4	US-09-538-092-918	Sequence 918, App
96	90	6.0	346	4	US-08-887-534A-68	Sequence 68, Appl
97	90	6.0	346	4	US-09-527-431-68	Sequence 68, Appl
98	90	6.0	346	4	US-09-446-861-68	Sequence 68, Appl
99	90	6.0	883	4	US-09-489-039A-12755	Sequence 12755, A
100	90	6.0	1427	4	US-09-538-092-1044	Sequence 1044, Ap

ALIGNMENTS

RESULT 1
US-08-557-115-3
; Sequence 3, Application US/08557115
; Patent No. 5731151
; GENERAL INFORMATION:
; APPLICANT: King, Harold C.
; APPLICANT: Sathish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysin
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,115
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 03063-0171US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 818-3700
; TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US-08-557-115-3

Query Match 91.1%; Score 1376; DB 1; Length 309;
Best Local Similarity 90.4%; Pred. No. 2.6e-132;
Matches 272; Conservative 16; Mismatches 13; Indels 0; Gaps 0;
QY 1 MTGIFAQTVVVKSAIETADGALDFYNNKYLDQVTPWKTFTDTIKELSRFKQEYSQASV 60
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVTPWQTFDTIKELSRFKQEYSQASV 60
QY 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120
DB 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKASAKDILIKVL 120
QY 121 DGVNKLNEAOKSLGSSQFNNSAGKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAG 180
DB 121 DGIITKLNKAOKSLVSSQFNNSAGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAG 180
QY 181 AAAGVAGPFGILISYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAAK 240
DB 181 AAAGVAGPFGILISYSIAAGVIEGKLIPELKNKLSVQNFFTLSNTVTKQANKDIDAAK 240
QY 241 LKLAETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTLLFV 300
DB 241 LKLTETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTLLFV 300

QY 301 P 301
DB 301 P 301
RESULT 2
PCT-US94-05869-3
; Sequence 3, Application PC/TUS9405869
; GENERAL INFORMATION:
; APPLICANT: King, C. H.
; APPLICANT: Sathish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIN
; TITLE OF INVENTION: OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 1200, The Candler Building
; STREET: 127 Peachtree Street, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05869
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Perlyman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US94-05869-3

Query Match 91.1%; Score 1376; DB 5; Length 309;
Best Local Similarity 90.4%; Pred. No. 2.6e-132;
Matches 272; Conservative 16; Mismatches 13; Indels 0; Gaps 0;
QY 1 MTGIFAQTVVVKSAIETADGALDFYNNKYLDQVTPWKTFTDTIKELSRFKQEYSQASV 60
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVTPWQTFDTIKELSRFKQEYSQASV 60
QY 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120
DB 61 LVGDIKVLLMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKASAKDILIKVL 120
QY 121 DGVNKLNEAOKSLGSSQFNNSAGKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAG 180
DB 121 DGIITKLNKAOKSLVSSQFNNSAGKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAG 180
QY 181 AAAGVAGPFGILISYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAAK 240
DB 181 AAAGVAGPFGILISYSIAAGVIEGKLIPELKNKLSVQNFFTLSNTVTKQANKDIDAAK 240
QY 241 LKLAETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTLLFV 300
DB 241 LKLTETIAAIGETETETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTLLFV 300
QY 301 P 301

STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3944:
SEQUENCE CHARACTERISTICS:
LENGTH: 1196 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1196
SEQUENCE DESCRIPTION: SEQ ID NO: 3944:
US-09-107-532A-3944
Query Match 7.2%; Score 108.5; DB 4; Length 1196;
Best Local Similarity 20.1%; Pred. No. 0.1;
Matches 66; Conservative 41; Mismatches 89; Indels 133; Gaps 12;
QY 3 GIFAQTVGVK-----SAETADG---ALDFVNYKLDQVVPKTFDETIKELSR 49
DB 162 GIP-ETAGVLKYKQKKAQKLPFTEDNLSRVQDIHELEEQLTPLAAQSEAAKEFLR 220
QY 50 FKQEYSQ-EASVLVGDIKVLLMDSQDKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKK 108
DB 221 LKETLTQTDVSLMVAEIKTAKD-----WDNKKQ 248
QY 109 ASAQKILIRILDDGVNKLNEA---QKSLGSSQSFNNASGKLALDSOLTNDFFSEKSY 165
DB 249 AQLAKFNL-----ELGKUSESIQESILAKQKENAQRDLIEKXQVQLDLSEKLLQ 302
QY 166 FQSQVDRIRKAYAGAAAGIVAGFFGLIISYSTAAGVIEGKLIPELNDRLKAVQNFFTSL 225
DB 303 TEQCKD-----VLQERTKTKTQSSQY 324
QY 226 SVTVKQAKNDIAAKLKLATEIAAIGETETTRFYVDYDMLSLKGAAK-----280
DB 325 QTSIAEAK-----KVK-----HFELQESLMKAAAEKETEIQ 357
QY 281 -----MINTCNEYQ---QRHGKKTLLVDP 302
DB 358 KAEANLIKTOQELEKYQKSTKELLAEIRD 386
RESULT 6
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIORITY FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080
Query Match 7.1%; Score 106.5; DB 3; Length 3696;
Best Local Similarity 21.5%; Pred. No. 1;
Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;
QY 11 EVKSAETADGALDFVNYKLDQVVPKTFDETIKELSR-----FKQEYSQ-EA 58
DB 568 EQVNDILIPSNYTLASVNY-----NKLKRAQTVLDEETNTFPNOKYSQTQI 616
QY 59 SVLVGDIKVLMM-----DSQDKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKKASA 111
DB 617 DLLLHELQTTLINRVSSASREINDKAOEMTDVYDSTELTTEKDT---LVDOJENHKNEI 673
QY 112 QKDILIRILDDGVNKLNEAOKSLGSS-----QSFNNASGKLALDSOLTNDFFS 160
DB 674 SNIDDELTDGVERVEKAGLHLESTPHVTPKPNARQVNNRA-----DQKTLIRN 727
QY 161 EKSSYFQSQVDRIRK-EAYAGAAAGIVAGFFGLIISYSTAAGVIEGKLIPELNDRLKAVQ 219
DB 728 NHEATTEQNEATQVEAHSSDA-----IAKIGEAETDTTVNE---ARD 768
QY 220 NFFTSLSVTVKQAKNDIAAKLKLATEIAAIGETETTRFYVDYDMLSLKGAAK 279
DB 769 NGTKLIATDVENPTKAAE---RAAVTNSANSKIKDINNNTQATLDERNDALVNRSKD 825
QY 280 KMINTCNEYQ-----QRHGKKTLLVDP 301
DB 826 EAIQNINTAQNDVDVTEAQNNGTNTIQVVP 855
RESULT 7
US-09-107-532A-5150
; Sequence 5150, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571

FILED DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Denke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5150:
SEQUENCE CHARACTERISTICS:
LENGTH: 1231 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...1231
SEQUENCE DESCRIPTION: SEQ ID NO: 5150:
US-09-107-532A-5150

Query Match 7.0%; Score 106; DB 4; Length 1231;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 77; Conservative 60; Mismatches 124; Indels 74; Gaps 17;

QY 6 AEQTVVVKSAIETADGALD-----FVKYLDQVIPWKTDE-----TIKELS 48
DB 269 AEEIVSGVSAQTIEQIIDYLNANGRTGFINHLHYRPPFVETVQKLPHTVKTIAVL 328
QY 49 RFKQBYSEASVVLGDIKVLMDSQ-----DKYFEATQTVYEMCGVVTQLLSAVILLF 101
DB 329 RSKBPAGCEPLL-DVQSALYDSLRPAVIGRYGLGSK-----DVTPOISA--VF 378
QY 102 DEYNEKASAKDILIRILDGYNKLENAQSL-LGSSQSFN-----NASGKL 148
DB 379 DELKQPSIRKRFITIGVDVVTYQSLPQKESLDLTPQTFOAKFWGFGSDGTVGANKSA 438
QY 149 LALDSQTNDFSEKSYFSQSQ-----VDRIK-----KEAYAGAAAGIVA--GPFGLII 194
DB 439 IKIIGDHTDKVAQGYFYDSKSGSLTVSHLRFGTPIRSAYLVHEDLVACHTP-AVILH 497
QY 195 SYSTAAGVIEGLIPELNDRKAVQNFSTLSVTVKQ--ANKDI--DAAKLKLATEIAA 249
DB 498 SYDLVKGLKPGGIF--LLNTLWSDQLETHPLPLKRYLAENNIRFYTNAMRLAQEVL 555
QY 250 IGEIKTETETTRF-----YVDVDDLMSLLKGAACK 280
DB 556 GRRINTAMETAPFKLADIIPFDE-VLPLLKGEALK 589

RESULT 8
US-09-328-352-4284
; Sequence 4284, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4284

Query Match 6.9%; Score 104; DB 4; Length 815;
Best Local Similarity 22.0%; Pred. No. 0.17;
Matches 72; Conservative 47; Mismatches 123; Indels 86; Gaps 15;

QY 9 TVEVVKSIAETADGALDFYNKYLDQVIPWKTDETIKELSRPKQ--EYSQSEASVLVGDIK 66
DB 220 TAHIIQSWADKADTEIIPQITILEPGLPYEPDVEIKAESEFKEPAEKDREGVRLDLA 279
QY 67 VLLMDSQDKYFEATQTVYEMCGVVTQLLSAVILLFDE--YNEKASAKDILIRI----- 119
DB 280 LVTIDGEDARD-----FDDAVYAEKRPGGYRVVVAIVASH 316
QY 120 --LDDGVNKLNEAKSLLGSSQSFNNASGKLALDSQLTNDFSEKSYFSQSOVDRI--- 173
DB 317 YVRILDSALNE--EAEER--GTSVVPFPH--VLPMLPEALSNGLCSLNPH----VDRLCMV 366
QY 174 --RKEAYAGAAAGIVAGPFG--IISYSIAAGVIEG--KLIPELNDRKAVQNFSTLS 225
DB 367 CDLKLRTGRVTGYEFYPAVMHSAKRLTYTQGVYFEGATDAIPKDRDIHKSILNLF-QL 425
QY 226 SVTVKQANKDIDAAKLKLATBIAAIGEIKTETETTRFYVDYDDL-----MLSLLKGAACK 280
DB 426 YQILKNLVRDRA-----MBFETIETMTFDELGGIKEILPRTREDAHK 469
QY 281 MINTC-----NEYQORHGKKTILLEV 300
DB 470 LIECMILLANVAAYALEHIDPMLYRV 497

RESULT 9

US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820
; LENGTH: 1211
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

Query Match 6.9%; Score 104; DB 3; Length 1211;
Best Local Similarity 18.0%; Pred. No. 0.31;
Matches 60; Conservative 60; Mismatches 131; Indels 82; Gaps 9;

QY 7 BOTVEVVK-----SAIETADGALDFYNKYLDQVIPWKTDETIKELSRFKQSEASV 60
DB 185 EESAGVLKYKKRKAESIQKLDHTEDNLNRVEDILYDLEGRVEPLKEAAIAKEYQLSKE 244
QY 61 LVGDIKVLMDSQDKYFEATQTVYEMC-----GVVTQLLSAVILLFDEYNE 106
DB 245 MEQSDVIVTVSDIDHYTQDNLRLNHLKSAQAEKQQAQINQLQKY----- 295
QY 107 KQASAKDILIRILDGYNKLENAQKSLGSSQSFNNASGKLALDSQLTND-----FS 160
DB 296 -KGKQQN-----DYDIEKLN---YELVAKATNEYQLSGKLVLEERKQSQSETNARYE 345
QY 161 EKSSYFQSOVDRIKKEAYAGAAAGIVAGPFGILLISYSIAAGVIEGKLIPELNDRL---K 216
DB 346 BELNLESQIDSIXNEKAQNE-----KLADLNKQKQLNK 381
QY 217 AVQNFSTLSVTVKQANKDIDAAKLKLATEIAAIGEIKTETETTRFYVDYDDLMSLLKG 276
DB 392 EVQELSLIYISDSQHDKEKLEIKNSYVTLNSEQSVVNDIRFLEHTINENEAKSRDLS 441
QY 277 AAKXMINTCNBYQQ-----RHGKKTILLEV 300

Wed Feb 2 11:26:40 2005

OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-09-710-279-1660

442 RLVEAFNLKDIQNNITQTOKEYQSSKSMKV 474

RESULT 10
US-09-134-000C-3594
Sequence 3594, Application US/09134000C
Patent No. 6617156

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3594
LENGTH: 284
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-3594

Query Match 6.8%; Score 103; DB 4; Length 284;
Best Local Similarity 21.0%; Pred. No. 0.039;
Matches 67; Conservative 46; Mismatches 106; Indels 100; Gaps 13;

QY 16 ATETADGALDFYNYKLDVVPKTFDETIKEL----SRFKQYSQEASVLDGDKVL--- 68
DB 6 AVKVGEGANFLEKNKEK-----VDYTIKFNDDATGLYKALENGEADAIVDDYPLGYA 59
QY 69 -----LMDSDQ-----KYFEATQTVVEMCGVVTQLLSAYILLFD 102
DB 60 VKNGQKQLQVGDKETGSSYFAVKGONPELIKFNAGLNKLDNGTYDKILNNYLATGD 119
QY 103 EYNEKASAKQDILIRILDGVNKLNEAKSLGSSQS-----FNASGKLALDSOLT 156
DB 120 ETNTQDAGEQ-----MKKTPKKEKYVIADSTFAPFQNAQGDYVGDVLDVL 168
QY 157 NDFSEKSSY-----FOSQVDRIRKEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLI 208
DB 169 KRAELQGFVVEKFGFSQVAVQAVE-----SQQADGWAG-----M 205
QY 209 PELNDRKA-----VQNFFTSLSVTVVQAN---KDIDAAK-----LKLATEIA-AIGEIKT 255
DB 206 TITDDRKKAFDFSPYVDFSGIQIAVKGNDKIKSYDDLKGGKVGKIGTESADFLKXKK 265
QY 256 ETETTRFYVDYDMLMLSL 274
DB 266 KYDSIKYLDITDLYSAL 284

RESULT 11
US-09-710-279-1660
Sequence 1660, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1660
LENGTH: 885
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

Query Match 6.8%; Score 103; DB 4; Length 885;
Best Local Similarity 18.2%; Pred. No. 0.24;
Matches 53; Conservative 52; Mismatches 111; Indels 76; Gaps 8;
QY 42 ETIKELSRFKQYSQEASVLDGDKVLLMDSQKYFEATQTVVEMC----- 87
DB 2 EPLKEFAATAKYSQKLSKEMEQSDVIVTVSDIDHYTEDNQRLDERNLHLSQQAESGQ 61
QY 88 GVVTVQLLSAYILLDFEYNEKASAKQDILIRILDGVNKLNEAKSLGSSQSFNNSGK 147
DB 62 AQINQLLQKY-----KGRQON-----DYDIEKLN---YELVKATENYQLSGK 102
QY 148 LLALDSQLTND-----FSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPFGLLIISYIAAG 201
DB 103 LNVLEERKKNQSETNARYEEELNLESQSDISIRKEAQN----- 142
QY 202 VIEGKLIPELNDRL---KAVQNFFTSLSVTVVQANKDIDAAKLKLAITEIAAIGEIKTET 257
DB 143 ----KLLAELKNKQKQNLKEVQELSLLYISDQHDKELEIKNSYITLMSQSDVNDI 198
QY 258 ETTRFYVDYDMLMLSLKGAAKMINTCNEYQ-----RHGKTKLLEV 300
DB 199 RFLHTINENEAKSRDLSDRLVEAFNLQDKIDQNNITQTOKEYQSSKSMKV 250

RESULT 12
US-09-914-259-66
Sequence 66, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 961
TYPE: PRT
ORGANISM: Bos taurus
US-09-914-259-66

Query Match 6.8%; Score 103; DB 4; Length 961;
Best Local Similarity 22.0%; Pred. No. 0.27;
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;

QY 11 EVVKSATETADGALDFYNYKVL-----DQVTPKTFDE-----TIKELSRFK 51
DB 639 BEVKTLQSHDSIVTHYKNMIREQDLQLEELKQIISTLKQNEQLQTAVTQVQSOIQHK 698
QY 52 QEYSQEASVLDGDKVLL-MDSQK--YFEATQTVVEMCGVVTQLLSAYILLDFEYNEKK 108
DB 699 DOYNL-----LKVQIGKDSQHGPTDGAQ-----MNGVQPEISR---LREEIBELK 743
QY 109 AS-----AQKQDILIRILDGVNKN--LNEAQSLSGSSQSFNNSGKLALDSOLTND 158
DB 744 SNRELLQSLAEKDSLIENLKSSQLSPGTNEQSSATAGDSEQIAELKQELATLSQL-NS 802
QY 159 FSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKL-----IPELN 212
DB 803 QSVETIKLQTEKQELLQKTEAFKAPVGESETVIATKTTD--VEGLSALLQETKELK 860
QY 213 DRKAVQNFFTSLSVTVVQANKDIDAAKLKLAITEIAAIGEIKTETTRFYVDYDMLMS 272
DB 861 NEIKALSEERTAIKQLDSSNSTI-----AILQNEKNKLEVDITDSKKEQDLDLV- 910

1313 LSSQLQDQELLQETROK-----LNVSTKLQLEBERNSLQDQLEMEAK 1359
QY 219 QNF---FTSLSVTVKQANKDIDAAKLKLAIEAIGIKTETETTRFYVDYDDLMLSLK 275
Db 1360 QNLERHISTNLQISDSKKLQ---DPASTVEALEEGK-----KRFQKEIENLTQYEE 1410
QY 276 GAA--KKMINTCNEYQQ 290
Db 1411 KAAAYDKLEKYNRLQQ 1427
RESULT 16
US-09-543-681A-8019
; Sequence 8019, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8019
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8019
Query Match 6.7%; Score 101; DB 4; Length 1129;
Best Local Similarity 23.8%; Pred. No. 0.56;
Matches 69; Conservative 48; Mismatches 128; Indels 42; Gaps 13;
QY 4 IFAEQTVVVKSAIEATGADGALDPYKYLQVTPWKTFTETIKELSRFKQYEQEASVLVG 63
Db 435 IADWLEENQTSINKAGSLDRAEIGDKA-----RNASQAERLGLI-KEADSNMG 487
QY 64 DIKVLMDSDQKYFEATQTVYEWGVVTL--LSAYILLFDEYNEKKAQKIDILRIID 121
Db 488 K-KSLMDDDINKY---QKIINERSAVEVELNSLAKLNKTQTRQATLORK--LQRIN 541
QY 122 DGVNKINEAKG--LLGSSOSFNASGKLLALDSQLTNDSEKSYFQSQVDRIRKEAVA 179
Db 542 DAENKLPALQSVLDLNPDKFRNEHRLRTRTANSUHRDRIQSAINRMTPLEREELDA 601
QY 180 GA---AAGIVAGPGLIISYIAAGVIEGK-----LIPE--LNDRLKAVQNF-- 222
Db 602 AADIVNKIIGAPSGIYVSELPDGLVKRAGFTKDRTLNIPDERIKDYLESVDVYVWYNY 661
QY 223 ----TSLSVTVKQANKDIDAAKLKLAIEAIGIKTETETTR 261
Db 662 IRQVAPEIELTAKGRVDMQ-NQIKAITAEYNQLADIATTPKRSR 706
RESULT 17
US-09-447-497-15
; Sequence 15, Application US/09447497
; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.
; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Susin, Santos A.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; PRIOR FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-447-497-14
Query Match 6.7%; Score 100.5; DB 4; Length 553;
Best Local Similarity 22.6%; Pred. No. 0.2;
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;
QY 17 IETADGALDPYKYLQVTPWKTFTETIKELSRFKQYEQEAS 59
Db 101 IVSEDPPELPMRPPLSKEL-WFSDDPNVTKTLRFQWNGKERSIYFQPPSFVSAQDLPH 159

EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 15
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human apoptosis-inducing factor (half isoform #2);
; OTHER INFORMATION: half-exon-skip-Cold; mature polypeptide
US-09-447-497-15
Query Match 6.7%; Score 100.5; DB 4; Length 526;
Best Local Similarity 22.6%; Pred. No. 0.19;
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;
QY 17 IETADGALDPYKYLQVTPWKTFTETIKELSRFKQYEQEAS 59
Db 74 IVSEDPPELPMRPPLSKEL-WFSDDPNVTKTLRFQWNGKERSIYFQPPSFVSAQDLPH 132
QY 60 ----VLVGDIDIKVLLMDSQD---KYFEATQTVYEWGVVVT---QLLSAYILLFDEYN 105
Db 133 IENGVAVLTKG-KVQVLDVRDNMVKLNDGSOITVEKCLIAATGTPRSLAI 183
QY 106 EKASQAQ---KDILIRILD--GVNKLNEAQKSL-----LGSSOSFNASG-KLLALD 152
Db 184 -DRAGAEVKSRTTLFRKIGDFRSLEKISREVKSITTIIGGFLGSELAC--ALGRKARALG 240
QY 153 SOLTNDSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPGLIISYIAAGVIE 204
Db 241 TEVIQLPPEKNGMKILPELSNWTKEVRE-----GVKMPNAIVQS---VGVS 289
QY 205 GKLIPELNDRLKAVQNF--TSLSVTVKQANKDIDAAKLKLAIEAIGIKTETETTRFYV 264
Db 290 GKLLIKLKDGRKV-----ETDHIVAAGVLEPNVELAKTGGLLEIDSDFGGRV 336
QY 265 D 265
Db 337 N 337
RESULT 18
US-09-447-497-14
; Sequence 14, Application US/09447497
; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.
; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Susin, Santos A.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; PRIOR FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-447-497-14
Query Match 6.7%; Score 100.5; DB 4; Length 553;
Best Local Similarity 22.6%; Pred. No. 0.2;
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;
QY 17 IETADGALDPYKYLQVTPWKTFTETIKELSRFKQYEQEAS 59
Db 101 IVSEDPPELPMRPPLSKEL-WFSDDPNVTKTLRFQWNGKERSIYFQPPSFVSAQDLPH 159

Qy 60 -----VLVGDIKVLLMDSQD---KYFEATQTVVEMCGVVT-----QLLSAYILLFDEYN 105
Db 160 IENGGAVALTGK-KVQVLDVRDNMVKLNDGSIYKCLATGTPRSLSAI-----210
Qy 106 EKKASQAQ---KDILIRILDD---GVNKLNEAKSL-----LGSSQSFNNASG-KLLALD 152
Db 211 -DRAGAEVKSRTTLFRKIGDFRSLEKISREVKSITIIIGGFLGSELAC--ALGRKARALG 267
Qy 153 SOLTNDSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIE 204
Db 268 TEVIQLPEKGMGKILPEYLSNWTMEKVRRE-----GVKMPNAIVQS-----VGVSS 316
Qy 205 GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLKLATEIAAIGETETETTRFYV 264
Db 317 GKLLIKLKDGRKV-----ETDHIHVAAGLEPNVELAKTGGLSDSDFGGRV 363
Qy 265 D 265
Db 364 N 364

RESULT 19

US-09-538-092-807
; Sequence 807, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormat Version 0.9
; SEQ ID NO 807
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number YPR070W
US-09-538-092-807

Query Match 6.7%; Score 100.5; DB 4; Length 566;
Best Local Similarity 25.0%; Pred. No. 0.21;
Matches 56; Conservative 29; Mismatches 92; Indels 47; Gaps 11;

Qy 15 SAITADGALDFYKQY-----LDQVTPWKTDFETI-KELSRFKQ-----52
Db 6 SYVETLDSMIELFDYKPGSITLENITRLCQTGLGESPTLELSNLSRLSTASKIIVIDV 65
Qy 53 EYSEASVLVGDIDKVLMLDSQD---YFEATQTVVEMCGVVTQLLSAYILLFDEYNEKKAS 110
Db 66 DYNNKQD-RIDQVKLVLASNFEDFYNQRDGEHEKSNILNSITKYPDLLKAFHNNLKL 124
Qy 111 AQDKILIRILDGYNKLN-EAQSLLGSSQSFNNASGKLLALDSQLTNDSEKSSY----165
Db 125 YLLDAYSHIESDTSIHNNGSSDKSLDSSNFSNN-QGKL-----DLFKYFTLSHYIROC 178
Qy 166 PQSQVDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLIP 209
Db 179 PQDNCCDFKVRTNLNDKFGI-----YILTQG-INGKEVP 211

RESULT 20

US-09-447-497-12
; Sequence 12, Application US/09447497

; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.
; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Susin, Bryan E.L.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; CURRENT FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human apoptosis-inducing factor (haIF isoform #1
; OTHER INFORMATION:); haIF-alt-exon-Gold; mature polypeptide
US-09-447-497-12

Query Match 6.7%; Score 100.5; DB 4; Length 609;
Best Local Similarity 22.6%; Pred. No. 0.24;
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16;

Qy 17 IETADGALDFYKQVLDQVTPWKTDFETIKLSRFSQYSQAS-----59
Db 157 IVSEDPPELPMRPPLSKEL-WFSDDPNVTTLRFPKQWNGKRSYTFQPPSYVSAQDLPH 215
Qy 60 -----VLVGDIKVLLMDSQD---KYFEATQTVVEMCGVVT-----OLLSAYILLFDEYN 105
Db 216 IENGGAVALTGK-KVQVLDVRDNMVKLNDGSIYKCLATGTPRSLSAI-----266
Qy 106 EKKASQAQ---KDILIRILDD---GVNKLNEAKSL-----LGSSQSFNNASG-KLLALD 152
Db 267 -DRAGAEVKSRTTLFRKIGDFRSLEKISREVKSITIIIGGFLGSELAC--ALGRKARALG 323
Qy 153 SOLTNDSEKSS-----YFQS-QVDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIE 204
Db 324 TEVIQLPEKGMGKILPEYLSNWTMEKVRRE-----GVKMPNAIVQS-----VGVSS 372
Qy 205 GKLIPELNDRLKAVQNFTSLSVTVKQAKDIDAAKLKLATEIAAIGETETETTRFYV 264
Db 373 GKLLIKLKDGRKV-----ETDHIHVAAGLEPNVELAKTGGLSDSDFGGRV 419
Qy 265 D 265
Db 420 N 420

RESULT 21

US-09-447-497-9
; Sequence 9, Application US/09447497
; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.
; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Susin, Bryan E.L.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; CURRENT FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0

230	IENG	GVAVLTGK-KVVQLDVRDNNVMVKLNDGSGQIYKELKIATGQTPRSLSAI	280
106	EK	SAQAQ---KDILIRILDD--GVNKLNEAQKSL-----LGSSQSFNNASG-KLLALD 152	
281	-DR	AGAEVKSRTTLFRKIGDPRSLEKISREVKSIITIIGGGFLGSELAC--ALGRKARALG 337	
153	SQ	TNDPFSKSS-----YFQS-QVDRIRKEAYAGAAGIVAGPFGILITSYSAAGVIE 204	
338	TE	VIQFLPPEKGNMGKILPEYLSNWTWEKVRRE-----GVKMPNAIVOS---GVSS 386	
205	GK	LPELNDLKAQVNFSTLSVTYKQANKIDDAKCLKLATEIAAIGEIKTETTTTFYV 264	
387	GK	LKILKQGRV-----ETDHIVAAGVLEPNVELAKTGGLSDSDFGGFRV 433	
265	D	265 D 265	
434	N	434 N 434	
RESULT 23			
US-09-447-497-8			
; Sequence 8, Application US/09447497			
; Patent No. 6773911			
; GENERAL INFORMATION:			
; APPLICANT: Penninger, Josef M.			
; APPLICANT: Kroemer, Guido P.			
; APPLICANT: Siderovski, David P.			
; APPLICANT: Zamzami, Naoufal			
; APPLICANT: Susin, Santos A.			
; APPLICANT: Snow, Bryan E.L.			
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR			
; FILE REFERENCE: 01017/36780			
; CURRENT APPLICATION NUMBER: US/09/447,497			
; CURRENT FILING DATE: 1999-11-23			
; EARLIER APPLICATION NUMBER: 60/109,595			
; EARLIER FILING DATE: 1998-11-23			
; NUMBER OF SEQ ID NOS: 25			
; SOFTWARE: Patent In Ver. 2.0			
; SEQ ID NO 8			
; LENGTH: 627			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-447-497-8			
Query Match 6.7%; Score 100.5; DB 4; Length 627;			
Best Local Similarity 22.6%; Pred. No. 0.25;			
Matches 68; Conservative 50; Mismatches 94; Indels 89; Gaps 16			
17	IB	TADGALDFYNNKYLQDVIPKWTDETIKELSRFKQEYSQEAS----- 59	
175	I	VSDPELPYRMPPLSKEL-WFSDDPNVTKLFRKQWNGKRSIYQPPSFYVSAQDLPH 233	
60	---	VLVGDIKVLLMSOD---KYFEATQTVYEMCGVVT---QLLSAYILLFDEYN 105	
234	I	ENGAVLTGK-KVVQLDVRDNNVMVKLNDGSGQIYKELKIATGQTPRSLSAI 284	
106	E	KASQAQ---KDILIRILDD--GVNKLNEAQKSL-----LGSSQSFNNASG-KLLALD 152	
285	-D	RAGAEVKSRTTLFRKIGDPRSLEKISREVKSIITIIGGGFLGSELAC--ALGRKARALG 341	
153	S	QTNDFSEKSS-----YFQS-QVDRIRKEAYAGAAGIVAGPFGILITSYSAAGVIE 204	
342	T	VIQFLPPEKGNMGKILPEYLSNWTWEKVRRE-----GVKMPNAIVOS---GVSS 390	
205	G	KLPELNDLKAQVNFSTLSVTYKQANKIDDAKCLKLATEIAAIGEIKTETTTTFYV 264	
391	G	KLILKQGRV-----ETDHIVAAGVLEPNVELAKTGGLSDSDFGGFRV 437	
265	D	265 D 265	
438	N	438 N 438	

RESULT 24
US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Weinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938.105
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wanneil M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 3595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-938-105-3

Query Match 6.6%; Score 100; DB 3; Length 1886;
Best Local Similarity 21.3%; Pred. No. 1.6; Indels 88; Gaps 12;
Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;
Qy 5 FAEQTVVVKSAIETADGALDFYNYKLDQVTPWKTFDTIKELSRFKQEQYSQEQASVLVGD 64
Db 1120 FQWRDLLEATLQHEATAALRKKHADV---AELGEQIDNLRQVKLEKEKS-----E 1172
Qy 65 IKVLLMSQDKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKKASAKQDILIRLDGV 124
Db 1173 FQWRDLLEATLQHEATAALRKKHADV---AELGEQIDNLRQVKLEKEKS-----E 1205
Qy 125 N-----KLINEAKSLGSSQSFNNASGKLLALDSQLTNDFSEKS-----SYFSQ 169
Db 1206 NEYRVKLEEAQRSL-----NDFTQRAKLQTEGELARQLEKEALIWLTGKLSYTOQM 1261
Qy 170 VDIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPE-----LND 214
Db 1262 EDLKRQLEEEGKAKNALA-----HALQSAHDCDLLREQYEEEMEAELQRLVLSKA 1313
Qy 215 LKAVONFTSLSVTVKQANKDIDAUKLATE-----IAAIGIKETETTT---RFVVD 265
Db 1314 NSEVAQWRKTYETDAIQRTBELEAKKLAQRLQDAEAEVAVNAKCGSSLEKTKHRLQNE 1373
Qy 266 YDDLMLSLKGAACKMINTCNEYQORHGKKTLL 299
Db 1374 IEDLMVDVVERSNA---AALDKKQORNFDKILAE 1404

RESULT 25
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751

; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310.187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1
Query Match 6.6%; Score 100; DB 3; Length 1939;
Best Local Similarity 21.3%; Pred. No. 1.7; Indels 88; Gaps 12;
Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;
Qy 5 FAEQTVVVKSAIETADGALDFYNYKLDQVTPWKTFDTIKELSRFKQEQYSQEQASVLVGD 64
Db 1173 FQWRDLLEATLQHEATAALRKKHADV---AELGEQIDNLRQVKLEKEKS-----E 1225
Qy 65 IKVLLMSQDKYFEATQTVVWCGVVTQLLSAYILLDFEYNEKKASAKQDILIRLDGV 124
Db 1226 FKLLEDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258
Qy 125 N-----KLINEAKSLGSSQSFNNASGKLLALDSQLTNDFSEKS-----SYFSQ 169
Db 1259 NEYRVKLEEAQRSL-----NDFTQRAKLQTEGELARQLEKEALISQTRGKLSYTOQM 1314
Qy 170 VDIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPE-----LND 214
Db 1315 EDLKRQLEEEGKAKNALA-----HALQSAHDCDLLREQYEEEMEAELQRLVLSKA 1366
Qy 215 LKAVONFTSLSVTVKQANKDIDAUKLATE-----IAAIGIKETETTT---RFVVD 265
Db 1367 NSEVAQWRKTYETDAIQRTBELEAKKLAQRLQDAEAEVAVNAKCGSSLEKTKHRLQNE 1426
Qy 266 YDDLMLSLKGAACKMINTCNEYQORHGKKTLL 299
Db 1427 IEDLMVDVVERSNA---AALDKKQORNFDKILAE 1457

RESULT 26
US-09-538-092-917
; Sequence 917, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538.092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 917
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P13533
US-09-538-092-917

Query Match 6.6%; Score 100; DB 4; Length 1939;

Qy	24	LDIFYNKYLDOVTPWKTFDETIKEISRPQEQYSQASVLVGDIKVLLMDSODKYFEATQTV	83
Db	159	LELYNEELCDLL---STDDTK--IRIPDDSTKKGSVIIQGLEBIPVHSKDDVYKLLKSG	213
Qy	84	YEWCGVVTQLLSAYILLFDEYNEKKASQAQDILIRILDG-----VNKLENAQKSL	135
Db	214	KERRKATATLWNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKILNV--DLA	263
Qy	136	GS---SOSTFN-----NASGKLALDSQLTN--DFSEKSYFSQSDVRIRKEAYAG	180
Db	264	GSENVSKAGNEKGIKRVRETVAINOSLTLGRVITALVDRAHPYVRESKLTLLQLQESLGG	323
Qy	181	AA-----AGIVAGPFGI-----ITSYSTAAGVIEGKLIPELNDRL--KAVQNPFTSLSVT	228
Db	324	RKTSIATISFGHKDIBETLSTLEYAHRKNIQNK--PEVNQKLTKTVLKYETE---E	378
Qy	229	VQKANKIDAAKLKLATRIA--ATGEIKTETETTRFYVDYDDLMLSLLK	275
Db	379	IDKLKRDLMARDKNGIVLABETTYGEITLKLESQNRRELNEKMLLLKALK	427

```

RESULT 30
US-10-270-085-8 :
; Sequence 8, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
; APPLICANT: Nislaw, Corey
; APPLICANT: Sakowicz, Ronan
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PrT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

```

RESULT 31
US-09-538-092-914
; Sequence 914, Application US/09538092

```

; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 914
; LENGTH: 2349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P12270
US-09-538-092-914

Query Match          6.8%; Score 99.5; DB 4; Length 2349;
Best Local Similarity 18.5%; Pred. No. 2.5;
Matches 60; Conservative 49; Mismatches 104; Indels 111; Gaps 11;

QY      27 YNKYL-DQVIPWKTDFETIKELSRPKOYSQ-----EASVL 61
DB      1354 YRKLLSEKVEHTKRIQQLTSEEIGRLKAEIARSNASLTNNQNLQSLKEDLNKVRTEKETI 1413
QY      62 VGDIKVLMLDSQDKYFEATQTVVGCWGVVTLISAYILLFDEYNKKKASQAQDKILIRILD 121
DB      1414 QKDLDAKIIDIQEKVITIQ-----VKKIGRRY---KTQTEELKAQODKVMETSAQS 1462
QY      122 DG-----VNKLNRAQSKLLGSSSFNNASGKLIALDSQLTN-----DFS 160
DB      1463 SGDHQEHQSVQEMQELKETL-----NQAETKSKLSQSVENLQKLTSEKETEARNLQ 1515
QY      161 EKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVON 220
DB      1516 EQTVQLQSELRSURQ-----DLQDRTTQEEQ 1541
QY      221 FFTSLSVTVKQANKIDAAKLIKATBIAAI-----GEIKTETETTFYVDYDDLMLS 272
DB      1542 LRQIITEKEBKTKAIVAAKSKIA-HLGVKQDLTKENEELKQRNGALDQKDQLDVIRIT 1600
QY      273 LLK-----GAAKMINTCNEYQQRH 292
DB      1601 ALKSQVEGRISRLERELREHQERH 1624

RESULT 32
US-09-134-000C-4522
; Sequence 4522, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4522
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; FEATURE:

```

Wed Feb 2 11:26:40 2005

NAME/KEY: MISC FEATURE
LOCATION: (8)_(8)
OTHER INFORMATION: Amino acid 8 is Xaa wherein Xaa = any amino acid.
US-09-134-000C-4522

Query Match
Best Local Similarity 6.5%; Score 98; DB 4; Length 618;
Matches 55; Conservative 51; Mismatches 104; Indels 54; Gaps 10;

91 TOLLSAYILLFD-EYNEKA---SAQKDL-IRILD----GVNKLNEAQSLLGSSQS 140
92 TQMSKOLVRDVAANLKALSDSNADISARVDDKGIIRATNDLN--QNNIGKND 129
141 FNN---ASGKLALDSQ-----LTNDFSK-----SSYFQSDVRIRKEAYAG 180
130 YRLNDFSTSKYQALNDKRVVNVVQPSGTGTGVLVYKSNLENKYQEITNTASIF 189
181 AAGVAGPFGGLIISYTAAGVIE-----GKLIPELNDRLKAVQNF 222
190 FTASIIAAAIISIVTLIARSITKPIGEMRQAIRIARGDVAGVYVHGKDELQLAETP 249
223 TSLSVTVKQANKDIDAAKLKLATEIAAICEIKTETETTRFYVDYDDLMLSLKGAACKMI 282
250 NQLSERIEBAQETMAERNLDSVLTHMTDGVIAIDRRGKVITINEMALSLLNVKNENVI 309
283 NT-----CNEYQQRHGKTKLLLEVPD 302
310 GTSLELLDIEDYTLRKLEEPD 333

RESULT 33
US-09-489-039A-8554
Sequence 8554, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489, 039A
PRIORITY FILING DATE: 2000-01-27
PRIORITY FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 8554
LENGTH: 1122
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8554

Query Match
Best Local Similarity 6.5%; Score 98; DB 4; Length 1122;
Matches 68; Conservative 57; Mismatches 109; Indels 96; Gaps 14;

8 QTVEVWKS---AIFTADGALDFYKYLDOVIPW-KTFDETIKELSRPKORYQASVLVG 63
57 ETVEVLQSAALNALEERKSLERARQYQDVIDNFKLQSLRAQLNLSSEPRQVPTGLTA 116
64 DIKVLMDSQDKYEATQTVYEWCGVVTOLLSAVILLFDEYNEKKASAKDKILIRILD-- 121
117 D-----ALNQELIQ-----VSSQLLE-----SSRQAEQDRAREIADSL 151
122 -----DGVNKLNEAQSLLGSSQSFNNA-----SGKILALDSQLTNDFSE 161
152 NQLPQQQTIDARRQLNEVERRI--GTQTGNALAAQNALQASARLKAIVDEL--DLAQ 207
162 KSSVFQSDVRIRKEAYAGAGVAGPFGGLIISYTAAGVIEGKLIPELNDRLKAVQNF 221
208 LSANNRQELSRARSE-----LAQKQSEQLDAYLQALNRL 241
222 FTLSVTVKQANKDIDAAKL-----KIATEIAAIGEIKTE-TETTRFYVDYDDLMLSL 274
242 --QNSQREAREKALESTELLAENSENLPDITAQFKVNRSLQALNQQAQRMDLVASQ 299

275 KGA-----AKKMINTCNEYQQRHGKTKLL 298
300 ROATNQTLOVRQALNTLREQSOWLGSSNLL 329

RESULT 34
US-09-540-236-2753
Sequence 2753, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 2753
LENGTH: 718
TYPE: PRT
ORGANISM: M. catarrhalis
US-09-540-236-2753

Query Match
Best Local Similarity 6.5%; Score 97.5; DB 4; Length 718;
Matches 66; Conservative 53; Mismatches 142; Indels 69; Gaps 11;

6 AEQTVVWKSJETAADGALDFYKYLDOVIPWKTETIKELSRFKQE-----YSQE 57
103 ATEKLEVAKEATQ-----DKVEKTSQSLVEDIKDKAQSLQEDAADTVLEAKQA 149
58 ASVLVGDIKV---LLMDSQDKYFEATQTVYEWCG--GVVTOLLSAYILLFDEYN-----E 106
150 ASDKVETTKARAQSLXDDATQTFESAKQAVEGKVEAIKEQVLDQVDSLKDQDNTDQD 209
107 KASAKDKILIRILDGVNKLNEAQSLLGSSQSF--NNASGKL-----LALD--SQTND 158
210 QEKQTLKQAVQAATAAKRKVEDVDVHTVFESFKNTASGKIDKQAAVDEKTEEVKSQ 269
159 FSEKSSYFQSDVRIRKEAYAGAGVAGPFGGLIISYTAAGVIEGKLIPELNDRLKAV 218
270 LSQKADALKSSGEELKQTAQTAANDAITAQ--AAVVGSGVAADSAQSTAQSAKDKL---- 325
219 QNFTLSVTVKQANKDIDAAKLKLATEIAAICEIKTETETTRFYVDYDDLMLSLKGA 278
326 -----NQLFEQKQSLDEKQVQELGKFGATEKINAVSENVDLATQVKEEA 371
279 KKMINTC-----NEYQQRHGKTKLL 297
372 QALQTNQAESLQAKAAGEEYDATHEDKGL 401

RESULT 35
US-08-785-428-2
Sequence 2, Application US/08785428
Patent No. 5750387
GENERAL INFORMATION:
APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5750387el tRNA Synthetase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

```

, SOFTWARE.: FastSEQ for Windows Version 2.0.
,
, CURRENT APPLICATION DATA:
,
, APPLICATION NUMBER: US/08/785,428
,
, FILING DATE: 17-JAN-1997
,
, CLASSIFICATION: 530
,
, PRIOR APPLICATION DATA:
,
, APPLICATION NUMBER: 9601096.2
,
, FILING DATE: 19-JAN-1996
,
, APPLICATION NUMBER: 9622617.0
,
, FILING DATE: 30-OCT-1996
,
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Gimmi, Edward R
,
, REGISTRATION NUMBER: 38,891
,
, REFERENCE/DOCKET NUMBER: P31354-5
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: 610-270-4478
,
, TELEFAX: 610-270-5090
,
, TELEX:
,
, INFORMATION FOR SEQ ID NO: 2:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 804 amino acids
,
, TYPE: amino acid
,
, STRANDEDNESS: single
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: peptide
,
, US-08-785-428-2

```

[illegible]

```

RESULT 36
US-08-996-797-2
; Sequence 2, Application US/08996797
; Patent No. 5994111
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5994111el tRNA Synthetase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia

```

STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,797
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/785,428
FILING DATE: 17-JAN-1997
APPLICATION NUMBER: 9601096.2
FILING DATE: 19-JAN-1996
APPLICATION NUMBER: 9622617.0
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31354-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:

Query Match	6.5%;	Score 97.5;	DB 2;	Length 804;				
Best Local Similarity	18.6%;	Pred. No. 0.75;						
Matches	69;	Conservative	63;	Mismatches 106; Indels 133; Gaps 19				
Qy	3	GIPAEQTVVVVKS	AIETADGALDPY	NKVLQDVIEMKTFDETIKLSRFBQBSO	EASVLV 62			
Db	82	GLPAEQ-----	YALDTGNDPRE	FTKQNI-----	QTFKQIKELG-PSYWDREVNVT-- 126			
Qy	63	GDIKVLMDSQDKY	FEATQTVY-----	EMGVVTVQLLSAV	ILFDFEYNE 106			
Db	127	-----	TDPEYKKTQWIF	IGLYNKGLAYVD	EVAVNWCALGTVLSNEEVI-DGVSE 176			
Qy	107	K-----	-----	KASAKQIILIRLP	-----	DCGVNKLNEAKSL	LGSSQ-----	SF--NN 143
Db	177	RGGHPVYRKMKQ	WVLKITEYADQL	LADLDDLPESL	KQMRNWI	GRSEGA	KVSFVDN 236	
Qy	144	ASGKL-----	-----	LALDSQVTNDF	SEKSSYFQSQ	VDNR--	IRKEAY 178	
Db	237	TEGKVEVFTTR	PDIY GASFLVLS	PEHALVNSIT	TDVEYKEKV	KAYQTEAS	KSDLERD 296	
Qy	179	AGAAAGIVAGP	FGI-----	-----	IISYSIA-----	AGVIEGKL	PELNDR-LKAVQNF 223	
Db	297	AKDKSGVFTG	AYAINPLSGE	KVQIWIADY	VLSTYGTGAI	-----	MAYPAHDDRDREYFAKKFDL 354	
Qy	224	SLSVTVKQANK	DIIDAAK	KLATEATA	AGE-----	-----	IKTETETTRFVVDYDMLMS 272	
Db	355	PIEIVEGGNV	-----	-----	EAAVYTGEGKH	INSGLDGL	LENAATK-----	AIQLL 398
Qy	273	LLKGAACK	QKWIN 283					
Db	399	EOKGAGEK	KVN 409					

RESULT 37
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5178
LENGTH: 1010
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 6.5%; Score 97.5; DB 3; Length 1010;
Best Local Similarity 18.6%; Pred. No. 1.1; Indels 129; Gaps 14;
Matches 66; Conservative 54; Mismatches 54

QY 38 KTFDETIKLSRF-----KQYSQASVVLGDIKVLIMDSQDK-----YFEATQTVYEW 86
DB 459 KLFQDLQDKSYLSKLEKEQLNEISSITNIDATLIDNDKDFVNEIKSAMSIGDT 518
QY 87 CGV-----VTQLLSAYILLFDEY----- 104
DB 519 CFIGCNEIHSIGEHIDFESIAQKNNKIKLESKKVIRDEIKIETRIEELNHNREINLF 578
QY 105 --NEKK--ASAQKDLIRLDGVDGNKLNKAEQKSLGSSQSFNNAS-----GKLLALDSOLT 156
DB 579 EKQEKDISELQKQL-----NHLNQLKDEQOSINKLVENFEKQKEIVNKIHQFDLDS 632
QY 157 -----NDFSEKSSY-----FQSQVDRIKQAYAGAAAGIVAGPFGLIISYI 198
DB 633 RNTQKQKLEIQINDPERSQFSSYNDPETYSHAKQ-----VETVEY 676
QY 199 AAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQANKDIDAOKLKLATEIAAIGETETE 258
DB 677 ENKTKDKL-NELNKKIKIEMNDQKHLTENLTQTSKEINNLEKKEKMQQLG----- 728
QY 259 TTRFYVDYDLDM-----LSLLKGAQKMINTCNEYQQRH-----GKKTL 297
DB 729 -----FESYDQVKSADLSAQKQEIERYINYNKYQSYEINRLKELVKGKLL 779

RESULT 38
US-09-914-259-19
Sequence 19, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: Fast-SEQ for Windows Version 4.0
SEQ ID NO 19
LENGTH: 975
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-914-259-19

Query Match 6.4%; Score 97; DB 4; Length 975;
Best Local Similarity 21.1%; Pred. No. 1.1;
Matches 63; Conservative 56; Mismatches 118; Indels 62; Gaps 14;
QY 37 WKTDFETIKLSRFRKQYSQASVVLGDIKVLIMDSQDKYFEATQTVYEWCGVVTQL--- 93

376 WRA-GETVKAEREQINMEDIMEAS-----TPNLEVEAAQTAATAAALAAQRTAL 422
QY 94 --LSAYILLFDEYNEK-KASAKDILIRLDGVDGNKLNKAEQKSLGSSQSFNNASGKLLA 150
DB 423 ANMSASVAV-----NEQARLATECERLYOQLDDKBEINQ-----QSYAEQLKEQVME 471
QY 151 LDSQLTNDFSEKSSYFQSQVDRIKQAYAG--AAGIVAGPFGLIISYIAAGVIEGKL- 207
DB 472 QBELIANARREYET-LQSEMARIQENESAKEVEKVLQALEELAVNYDOKSQEIDKNK 530
QY 208 -IPELNDRLKAVQNFSTLSVTVKQANKDIDAOKLKLATEIAA-----IGEI----- 253
DB 531 DIDALNEELQKQSVFNAASTELOQL-KDMSHQKRRITEMTJNLRLDLGEVGOAIPGE 589
QY 254 -----KTETETTRFYVDYD--DLMLSLKGAQKMINTCN--EYQQRHKGKTKTLE 299
DB 590 SSIDLKMSALAGTDASKVEEDFTMARLFISKMTAKNIAQRCNSMNETQOADSNNKISE 648

RESULT 39
US-09-248-796A-19212
Sequence 19212, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 19212
LENGTH: 317
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-19212

Query Match 6.4%; Score 96.5; DB 4; Length 317;
Best Local Similarity 19.1%; Pred. No. 0.22;
Matches 53; Conservative 53; Mismatches 95; Indels 77; Gaps 10;
QY 8 QTVWVKSAJETADGALDFYNKYLDQVIPWKTDFETIKLSRFRKQYSQASVVLGDIKV 67
DB 34 QKISAIKEKLE-----MDSN-----DESSRKNLQYISSYNRKLETKJRELKV 76
QY 68 L---LMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKQ-----DILIR 118
DB 77 MQERLTKEHDDYROEK-----ALLAKLYDIDFLYNETQEQVANUTKYAETDLSILVK 129
QY 119 ILDDGVNKLNEAQKSLGSSQSFNNASGKLLALDSOLTNDPSEKSSYFQSQVDRIKQAY 178
DB 130 ITEDLSNDLSMQTSYKNLEKSYQNWG--ITIDSQVNEINLK-----DLLKLAED-- 178
QY 179 AGAAAGIVAGPFGLIISYIAAGVIEGKLIP-----ELNDRKAVQNFST 224
DB 179 -----ITRNKQKQIDGKILKNDQNDDELTTKQMELEGQILIVODQLNK 223
QY 225 LS-----VTVKQANKDIDAOKLKLATEIAAIGETETE 258
DB 224 LNRKYQDLIVQADKESAVKLEETKQEFNDILSEKE 261

RESULT 40
US-09-248-796A-19055
Sequence 19055, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

;; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 107196.132
;; CURRENT APPLICATION NUMBER: US/09/248,796A
;; PRIOR FILING DATE: 1999-02-12
;; PRIOR APPLICATION NUMBER: US 60/074,725
;; PRIOR FILING DATE: 1998-02-13
;; PRIOR APPLICATION NUMBER: US 60/096,409
;; PRIOR FILING DATE: 1998-08-13
;; NUMBER OF SEQ ID NOS: 28208
;; SEQ ID NO 19055
;; LENGTH: 858
;; TYPE: PRT
;; ORGANISM: Candida albicans
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: (851)
;; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown

US-09-248-796A-19055

Query Match 6.4%; Score 96.5; DB 4; Length 858;
Best Local Similarity 20.0%; Pred. No. 1;
Matches 56; Conservative 46; Mismatches 107; Indels 71; Gaps 9;
US-09-248-796A-19055

Qy 8 QTEVVKSAIETADGALDFYNYK--LDQVVPKTFDETIKELSRFKQEQYSQEAASVILGDI 65
Db 489 QLVQEQINESIRIKNNFQIMSKNFQRLQSLSEKENDKNLDFSKNNHQOQOQOQIQQLLEQ 548
Qy 66 KV-----LMDSDQKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASAKQDILI 117
Db 549 KLIYEKICITILQDELQYKQPSDTTNTNNNNNNNNSSSYNNHNRSSLSNELNL 608
Qy 118 -----RIILDDGVNKLNEAQKSLGSS--QSFNNASGKLLALDLSQLTNDFSEKSS 164
Db 609 VNDYLQQLQTSYRINDELQVNDYK--LLNSSTLEKINN-----LTSKLQEKSI 656
Qy 165 YFQSQVDRIRKEAYAGAAAGIVAGPFGLLISYSTAAGVIEGKLPELNDRLKAVQNFT- 223
Db 657 ELRNQMG-----LNNKLQTELNLSDLKQRYNTE 685
Qy 224 --SLSVTVVKQANKDIDAAKLKL--ATEIAAIGEIKTETET 259.
Db 686 RIQMSYTVDSLRKNEALQKLVNKLTLDTMTDRFAESNS 725

RESULT 41
US-08-592-126-148
; Sequence 148 Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; NUMBER OF SEQUENCES: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 324-0880
;; TELEFAX: (415) 324-0960
;; INFORMATION FOR SEQ ID NO: 148:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1312 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-08-592-126-148

Query Match 6.4%; Score 96.5; DB 2; Length 1312;
Best Local Similarity 18.7%; Pred. No. 2.1;
Matches 74; Conservative 54; Mismatches 99; Indels 169; Gaps 16;

Qy 10 VEVVKSIAETADGALDFYNYKLDQVVPKTFDETIK-----ELSRFKQEQY 54
Db 502 MEVISLQNEKAD--LDRTLRLKLDQEMEQLNHHHTTTTQMEMELTKDKADKQEQIRKIKSRH 559
Qy 55 SQEASVLVGDIKVLMDSDQKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASAKQD 114
Db 560 SDELTSLIG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 591
Qy 115 ILIRILDDGVNKLNEAQKSLGSSQSFNNASGKLLALDLSQLTNDFSEK-----SSYFQS 168
Db 592 -----DRLAKLN--KELASSEQNNHNNELKRREEQLSS-YEDKLFVCGSQDPES 640
Qy 169 QVDRIRKE-----AYAGAAAGIVAGPFG----- 190
Db 641 DLRLKEIEKSKQRAMLGATA--VYSQFITQLTDENQSCCPVCQVRFQTEAELQEV 698
Qy 191 -----GLI--ISYSTAAGVIEGKLPELNDRLKA 217
Db 699 SDLOSRLRLAPDKLSTESLKKKEKRDDEMLGLVPMRQSIID--LKEKEIPELRNKLQN 756
Qy 218 VQNFFTSLSVTVVKQANKDIDAAKLKLATEIAAIGEIKTETET-----RFYVDY 266
Db 757 V-----NRDIQRLKNDIEQEETLLGTIMPEESAKVCLTDVIMERFQNEL 802
Qy 267 DDLMLSLKGAAG-----KMINTCNEYQORHKK 295
Db 803 KDVVERKIAQQAQKLGIDLDRTVQVQVNAQKQEKQHK 838

RESULT 42
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126

FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 6.4%; Score 96.5; DB 2; Length 1312;
Best Local Similarity 18.7%; Pred. No. 2.1;
Matches 74; Conservative 54; Mismatches 99; Indels 169; Gaps 16;
QY 10 VEVVKSATADGALDFYNYKLDQVTPKTFDETIK-----ELSRFKQY 54
Db 502 MEVISLQNEKAD--LDRTLKLDQMEQLNHHHTTTRTQMEMLTKDKADKDEQIRKIKSRH 559
QY 55 SQASVLVGDIVKLLMDSQDKYFEATQTVYVCGVVTQLLSAYILLDFEYNEKKASAKQD 114
Db 560 SDELTSLG-----YFPNKKQLEDMLHKS-----EINQTR----- 591
QY 115 ILIRILDGYNKLENAQKSLGSSQFNNAAGKLLALDSQLTDFSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKNHNNELKKEEQLSS--YEDKLFVDCGSDQFES 640
QY 169 QVDRIRK-----AYAGAAGIVAGPF----- 190
Db 641 DLRLKEEIEKSKORAMLAGATA--VYSOFITQLTDENQSCCPVQCFVTEAELOEVI 698
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNDRLKA 217
Db 699 SDLSQKLRLAPDKLSTESLKKKRRDEMGLVPMRQSIID--LKEKEIPELNRKLN 756
QY 218 VONFFTSLSVTKQANKDIDAALKLATEIAAIGEIKTETET-----RFVVDY 266
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDTVIMERFQME 802
QY 267 DDLMLSLKGA-----KMINTCNEYQORHGKK 295
Db 803 KDVVERKIAQAQAKLOGIDLDRTVQOVNQEKEQKH 838

RESULT 43
US-09-168-595-148
Sequence 148, Application US/09168595
Patent No. 655566
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/168,595
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,126
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-09-168-595-148

Query Match 6.4%; Score 96.5; DB 4; Length 1312;
Best Local Similarity 18.7%; Pred. No. 2.1;
Matches 74; Conservative 54; Mismatches 99; Indels 169; Gaps 16;
QY 10 VEVVKSATADGALDFYNYKLDQVTPKTFDETIK-----ELSRFKQY 54
Db 502 MEVISLQNEKAD--LDRTLKLDQMEQLNHHHTTTRTQMEMLTKDKADKDEQIRKIKSRH 559
QY 55 SQASVLVGDIVKLLMDSQDKYFEATQTVYVCGVVTQLLSAYILLDFEYNEKKASAKQD 114
Db 560 SDELTSLG-----YFPNKKQLEDMLHKS-----EINQTR----- 591
QY 115 ILIRILDGYNKLENAQKSLGSSQFNNAAGKLLALDSQLTDFSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKNHNNELKKEEQLSS--YEDKLFVDCGSDQFES 640
QY 169 QVDRIRK-----AYAGAAGIVAGPF----- 190
Db 641 DLRLKEEIEKSKORAMLAGATA--VYSOFITQLTDENQSCCPVQCFVTEAELOEVI 698
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNDRLKA 217
Db 699 SDLSQKLRLAPDKLSTESLKKKRRDEMGLVPMRQSIID--LKEKEIPELNRKLN 756
QY 218 VONFFTSLSVTKQANKDIDAALKLATEIAAIGEIKTETET-----RFVVDY 266
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDTVIMERFQME 802
QY 267 DDLMLSLKGA-----KMINTCNEYQORHGKK 295
Db 803 KDVVERKIAQAQAKLOGIDLDRTVQOVNQEKEQKH 838

RESULT 44
US-09-248-796A-18018
Sequence 18018, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725

```

: PRIOR FILING DATE: 1998-02-13
: PRIOR APPLICATION NUMBER: US 60/096,409
:
: PRIOR FILING DATE: 1998-08-13
: NUMBER OF SEQ ID NOS: 28208
:
: SEQ ID NO 18018
: LENGTH: 360
: TYPE: PRT
:
: ORGANISM: Candida albicans
: US-09-248-796A-18018

```

Query Match	6.4%;	Score 96;	DB 4;	Length 360;
Best Local Similarity	24.6%;	Pred. No. 0.3;		
Matches 46;	Conservative 32;	Mismatches 63;	Indels 46;	Gaps 8
Qy	8	QTVEVVKSAJETADGAL----	DFNKKYLDQVPIPKWTFDET	IKELSRFKQEVSEASVLVG 63
Db	54	QTIHEVQAVASTDKATEEKDDSLK	QDQFLAKHKVHRHKLQL-----	EAE----- 100
Qy	64	DIKVLMDSQDYFEATQTVVEMCG	VTVQLLSAYILLDFEYNEKKAS---	AQOKILIRIL 120
Db	101	--EPLLVENKRRYYWFFIRYHE-----	IWNFKYKAAEASFWEIDLSKDL	144
Qy	121	DDGWNKLEAQSLLGSSQSFNNA	SGKLLADSLTDNDFS-----	EKSSY-FOSQVDR 172
Db	145	DDWNNKLNERNYFISRVLAIFA	ASDGIVG--ENLIENFSTEVQL	PEAKSPYGFQIMMEN 202
Qy	173	IRKEAYA	179	
Db	203	IHSETY	209	

```

RESULT 45
US-09-543-681A-7696
; Sequence 7696, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7696
; LENGTH: 1086
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7696

```

Query Match	6.4%;	Score 96;	DB 4;	Length 1086;
Best Local Similarity	19.8%;	Pred. No. 1.7;		
Matches 48;	Conservative 56;	Mismatches 76;	Indels 62;	Gaps 12
Qy	88	GVVTQLLSAYILLPDEYNEKKAQAQKQIL-----IRILDDGVNKLNEAKQSILGSSQSFN	142	
Db	421	GIATGLLTAGAWAYEYQQQAQEQSRQKALEYAESLDVATESLNKLSDAQRQ-----	472	
Qy	143	NASGKLI-ALDSQLTNDFSEKSSVYFQSQVDRIRKEAYAGAAGIVAGFPGLIISY-----	196	
Db	473	-SIGLKAEGDQAQI-----EKIEELKRQOEOLKSSDLGGANK-----GISFYTGTGYE	520	
Qy	197	SIAAGVIEGKLIPELNDRLKAVQNFFTLSVTVVQANKDIDAAKILKLATEAAGIEIKTE	256	
Db	521	SFSAEYVVKAR-----KQELQVTDLETK-----NRELEVTRKMMALLDA--EVQKN	565	
Qy	257	TETRFYVDY-----DDLMLSI-LKGAA-----KKMINTCNEYQOBHGKKTLLFV	300	
Db	566	GMWSNALVYASRIYDASVKDEGLIKUHGQSAFDNLQAQIRSTAQEQQOQFSAQSLVI	625	
Qy	301	PD 302		

DB 626 SD 627

```

RESULT 46
US-09-538-092-1077
  / Sequence 1077, Application US/09538092
  / Patent No. 6753314
  / GENERAL INFORMATION:
  /   APPLICANT: Glact, Loic
  /   APPLICANT: Manfield, Traci A.
  /   TITLE OF INVENTION: Protein-Protein C
  /   FILE REFERENCE: 15966-542
  /   CURRENT APPLICATION NUMBER: US/09/538
  /   CURRENT FILING DATE: 2000-03-29
  /   PRIOR APPLICATION NUMBER: 60/127,352
  /   PRIOR FILING DATE: 1999-04-01
  /   PRIOR APPLICATION NUMBER: 60/178,965
  /   PRIOR FILING DATE: 2000-02-01
  /   NUMBER OF SEQ ID NOS: 1387
  / SOFTWARE: CuraPatSeqformatter Version
  / SEQ ID NO 1077
  /   LENGTH: 1960
  /   TYPE: prt
  /   ORGANISM: Homo sapiens
  /   FEATURE:
  /     NAME/KEY: misc_feature
  /     LOCATION: (0)...(0)
  /     - OTHER INFORMATION: Polypeptide Access
US-09-538-092-1077

```

[illegible]

RESULT 47
US-09-538-092-936
; Sequence 936. Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965

```

Best Local Similarity 19.2%; Pred. No. 4.9;
Matches 61; Conservative 52; Mismatches 101; Indels 103; Gaps 13;

2y 7 EQTVVVKSAIETA-----DGAIDFYNKYLDQVIPWKTFDTIKELSRPK 51
Db 1173 EQEVTYVKALDEETRSHEAQVQEMRQKHQTQAVELTEQLEQ-----FKRAKANLDKSK 1226

2y 52 QEYSQASVVLGDIVKLLMSQDKYFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
Db 1227 QTLKEKNADLAGELRVL-----GQAKQEV-----EHHKKKLEV 1259

2y 112 Q-KDILIRI-----LDDGVNKLNEAOKSLLGSSQSFNNASGKLALDLSOUTNDFSE 161
Db 1260 QLQDLQSKCSDGERARAEISDKVHKLQNEVESVTG---MLNEABGKAI---KLAKDVAS 1312

2y 162 KSYFQSGVDRIKRYAGAAAGIVGPFGLIISYSAAGVIE---GKLIPELNDRLKAV 218
Db 1313 LGSQLQDTQELLQOETRQK-----LNVSTKRLQLEDERSNLQDLDENEAK 1359

2y 219 QNF---FTSLSVTVKQANKDI-----DAAKKLATIAAIGE-----IKTE 256
Db 1360 QNLERHVSTINLTQSDSKKLQDFASTIEVMEEGKRLQKEMEGLSQQVEEKAAYDKLE 1419

2y 257 TETTRFYVDYDMLMSL 273
Db 1420 KTKNRLQOELDDLVDL 1436

```

RESULT 49

US-08-434-000A-4
; Sequence 4, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 746 amino acids
;

```

; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 936
; LENGTH: 2871
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P15924
US-09-538-092-936

      6.4%; Score 96; DB 4; Length 2871;
Query Match      17.6%; Pred. No. 8;
Best Local Similarity
Matches 59; Conservative 56; Mismatches 92; Indels 128; Gaps 12;

QY      28 NKYLQVIPWKTFDETIKELSRKQESQEASVLVG-----DIKVLMDSQDKYFEAT 80
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1208 NKYTEINITKT--TIKEISMQKEDDSKLNRLQDLRLSRENRLDKDEIVRLNLSILOAT 1264

QY      81 Q-----TVYEWGVVTQLLSAYILLFPEYNEKASQAQ---KDLIRILDGV---N 125
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1265 EQRRAEENALQOKACG-----SEIMQKKQHLEIEUKVQWQORSEDNAHHQ 1311

QY      126 KLNKAAQKSLGSGSQGF-----NNASGKLALDS----- 153
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1312 SLEAAKTTQDKNKEIERLKAQFEAKRRWEYENELSKVRNNYDEEIIILKNQFETIN 1371

QY      154 -----QLTNDFSEKSSYFQGVDRIRKEAVAGAGIAGVPGGLIISYSIAAGVIEGK 206
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1372 IRTKTIHQLTMQKEEDTSYRAQIDNLTRNRS----- 1404

QY      207 LIFELNDRLKAVGNFFTSLSVTVKQAKNDIDAAKL-----KLATEIAAIGEIKTE 256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1405 -----LSEETIRLKNLTUTQTTENLRVVEEDIQOKATGSEVSRQKQLEVELRVQTMRT 1460

QY      257 TETTRFYVDYDDIMLSLLKGAAGKGMINTNEVQQR 291
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1461 -ESVRYKOSLDD-----AAKTIQDKNKEIER 1485

```

RESULT 48

US-08-875-435B-3
; Sequence 3, Application US/08875435B
; Patent No. 659304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-875-435B-3
Query Match 6.3%; Score 95.5; DB 4; Length 1972;

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Human Polyimmunoglobulin Receptor
US-08-434-000A-4

Query Match      6.3%; Score 94.5; DB 3; Length 746;
Best Local Similarity 20.7%; Pred. No. 1.3;
Matches 41; Conservative 47; Mismatches 69; Indels 41; Gaps 11;

QY 86 WCGV-----VTQLLSAYILLFDEYNEKKASAKDILI-----RILDDGVNKLNEAQ 131
Db 525 WCGVKQGHFYGETAAVYVAV-----EERKAAGSRDVS LAKADAAPDEKVLDSG---FREIE 577
QY 132 KSLIGSSQSFNNASCKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAGAAAGIVAG--P 189
Db 578 NKAIQDPRLP--ABEKAVA-----DTRDAQGSRASVDSSGSEEGGSSRALVSTLVP 628
QY 190 FGLIISY-SIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQ-----ANKDIDAALK 242
Db 629 LGLVLAVGAVGVARARHRKNV-DRV-SIRSYRTDISMSDFENSREFGANDNMGCASSIT 686
QY 243 LATEIAAIGEIKTETETT 260
Db 687 QETSLGGKEEFVATTEST 704

RESULT 50
US-09-312-157-4
; Sequence 4, Application US/09312157
; Patent No. 6303341
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/312,157
; FILING DATE: 14-May-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/434,000
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-351
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 746 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
;
; DESCRIPTION: Human Polyimmunoglobulin Receptor
; SEQUENCE DESCRIPTION: SEQ ID NO: 4
US-09-312-157-4

Query Match      6.3%; Score 94.5; DB 3; Length 746;
Best Local Similarity 20.7%; Pred. No. 1.3;
Matches 41; Conservative 47; Mismatches 69; Indels 41; Gaps 11;

QY 86 WCGV-----VTQLLSAYILLFDEYNEKKASAKDILI-----RILDDGVNKLNEAQ 131
Db 525 WCGVKQGHFYGETAAVYVAV-----EERKAAGSRDVS LAKADAAPDEKVLDSG---FREIE 577
QY 132 KSLIGSSQSFNNASCKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAGAAAGIVAG--P 189
Db 578 NKAIQDPRLP--ABEKAVA-----DTRDAQGSRASVDSSGSEEGGSSRALVSTLVP 628
QY 190 FGLIISY-SIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQ-----ANKDIDAALK 242
Db 629 LGLVLAVGAVGVARARHRKNV-DRV-SIRSYRTDISMSDFENSREFGANDNMGCASSIT 686
QY 243 LATEIAAIGEIKTETETT 260
Db 687 QETSLGGKEEFVATTEST 704

Search completed: January 28, 2005, 19:17:33
Job time : 43 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:03:34 ; Search time 157 Seconds
(without alignments)
692.325 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIFAEQTVEVKSARIETA.....TCNEYQORHGKKTLLVDPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : A Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1466	97.1	305	6	ABG73265
2	1466	97.1	749	6	ABG73266
3	1370	90.7	309	2	AAR67645
4	1065	70.5	1035	4	ABG25221
5	126.5	8.4	582	6	ABM68103
6	116	7.7	941	6	ABU18419
7	113	7.5	1015	6	ABP76825
8	113	7.5	1015	6	ABP78617
9	111.5	7.4	1036	7	ADF07451
10	110.5	7.3	2681	6	ABJ19025
11	110	7.3	1189	6	ABU42996
12	109	7.2	515	7	ADC95690
13	108.5	7.2	1196	7	ADC94317
14	107.5	7.1	742	6	ABP25599
15	106.5	7.1	3692	6	ABU43311
16	106.5	7.1	3696	5	ABP40235
17	106	7.0	1231	7	ADC95823
18	106	7.0	3187	7	ADC956031
19	106	7.0	3187	7	ADC956035
20	105.5	7.0	588	4	ABJ1280
21	104.5	6.9	1292	5	ABJ77986
22	104	6.9	685	4	ABJ64018
23	104	6.9	722	4	AAU35203
24	104	6.9	815	6	ADA32997
25	104	6.9	1211	5	ABP39975

26	104	6.9	2334	5	ABG31849
27	103.5	6.9	541	5	ABP29330
28	103.5	6.9	660	4	AAU38921
29	103.5	6.9	660	7	ADD42726
30	103.5	6.9	1020	6	ABU30092
31	103.5	6.9	1163	6	ABU23940
32	103.5	6.9	1976	7	ADE63514
33	103.5	6.9	1976	7	ADE63518
34	103.5	6.9	5373	4	AAU14603
35	103.5	6.9	5373	7	ADJ68935
36	103	6.8	284	7	ADH85709
37	103	6.8	885	4	AAH85283
38	103	6.8	961	7	ADB67135
39	103	6.8	961	8	ADL99406
40	103	6.8	1583	7	ADJ69829
41	103	6.8	2056	4	ABB59344
42	102.5	6.8	472	6	ABU27359
43	102.5	6.8	746	3	AAH46982
44	102.5	6.8	788	3	AAH46981
45	102.5	6.8	5447	4	AAU14697
46	102	6.8	590	4	AAE10039
47	102	6.8	590	4	AAU27604
48	102	6.8	612	4	AAE10042
49	102	6.8	612	4	AAU27607
50	102	6.8	739	4	AAH96493
51	102	6.8	1948	5	ABP73774
52	101.5	6.7	476	2	AAH43563
53	101.5	6.7	883	6	ADB10082
54	101.5	6.7	890	6	ADB10080
55	101.5	6.7	1857	5	AAU84350
56	101.5	6.7	1938	6	ABR92127
57	101.5	6.7	1945	6	ABU10398
58	101.5	6.7	1972	6	ABR92126
59	101.5	6.7	1972	7	ADE58724
60	101.5	6.7	1972	7	ADE58721
61	101.5	6.7	1972	7	ADE63520
62	101.5	6.7	1972	7	ADE63516
63	101.5	6.7	1972	8	ADN03745
64	101.5	6.7	1972	8	ADN03745
65	101.5	6.7	1979	6	ABU10399
66	101	6.7	1129	7	ADF07734
67	100.5	6.7	425	8	ADG98222
68	100.5	6.7	472	2	AAH37727
69	100.5	6.7	529	6	ABU20414
70	100.5	6.7	553	3	AAH97403
71	100.5	6.7	566	6	ABR53540
72	100.5	6.7	566	7	ADK62550
73	100.5	6.7	609	7	ADJ70327
74	100.5	6.7	609	7	ADJ68760
75	100.5	6.7	613	3	AAH15566
76	100.5	6.7	613	3	AAH15558
77	100.5	6.7	613	3	AAH92512
78	100.5	6.7	613	5	ABH89626
79	100.5	6.7	613	5	ABD25643
80	100.5	6.7	613	7	ADE48157
81	100.5	6.7	613	7	ADJ68765
82	100.5	6.7	623	3	AAH97401
83	100.5	6.7	627	3	AAH97399
84	100.5	6.7	803	4	AAU34101
85	100.5	6.7	805	4	AAU36693
86	100.5	6.7	861	8	AAH48067
87	100.5	6.7	1009	6	ABU15864
88	100.5	6.7	1377	4	ABB65439
89	100.5	6.7	7201	4	ABB71136
90	100	6.6	490	8	ADM72386
91	100	6.6	523	5	ABB78800
92	100	6.6	927	5	ABB49720
93	100	6.6	990	7	ADM05211
94	100	6.6	1174	6	ADA54343
95	100	6.6	1886	2	AAW54241
96	100	6.6	1939	5	ABB77096
97	100	6.6	2481	5	ABG92080
98	99.5	6.6	610	5	ABG80347

Abg31849	Human kin
Abp29330	Streptococ
Aau38921	C. tracho
Add42726	Chlamydia
Abu30092	Protein e
Abu23940	Protein e
Ad663514	Rat Prote
Ad663518	Rat Prote
Aau14603	Novel bon
Adj68935	Human hea
Adh85709	Enterococ
Adg82283	S. epide
Adb67135	General v
Adl99406	Nanostruc
Adj69829	Human hea
Abb59344	Drosophil
Abu27359	Protein e
Agg46982	Arabidops
Aag46981	Arabidops
Aau14697	Novel bon
Aae10039	N. mening
Aau27604	Neisseria
Aae10042	N. mening
Aau27607	Neisseria
Aab96493	Putative
Abp73774	Candida a
Aar43563	Hyalurona
Abi10082	Alloiooc
Abi10080	Alloiooc
Aau84350	Protein M
AbR92127	Human cer
Abu10398	Human em
AbR92126	Human cer
Ad58724	Human Pro
Ad58721	Human Pro
Ad663520	Human Pro
Ad663516	Human Pro
Adk70527	Respirato
Adn03745	Antipsoi
Abu10399	Human em
Adf07734	Bacterial
Adg98222	Human apo
Aay37727	Protein i
Abu20414	Protein e
Aay97403	Human apo
AbR53540	Protein s
Adk62550	Disease t
Adj70327	Human hea
Adj68760	Human hea
Aab15566	Fragment
Aab15558	Apoptosis
Aay92512	Human OXR
Abh89626	Human pol
Abd25643	Binding d
Ad648157	Human AIF
Adj68765	Human hea
Aay97401	Human apo
Aay97399	Human apo
Aau34101	Staphyloc
Adm48067	Staphyloc
Abu15864	Protein e
Abb65439	Drosophil
Abb71136	Drosophil
Adm72386	A. thalia
Abd78800	Human NOV
Abb49720	Listeria
Adm05211	Human pro
Ada54343	Human pro
Aaw54241	Rattus no
Abb77096	Human alp
Abg92080	Human rec
Abg80347	Clostridi

99 99.5 6.6 1048 4 ABB59245 Abb59245 Drosophil
100 99.5 6.6 1066 4 AAG67418 Aag67418 Amino aci

ALIGNMENTS

RESULT 1
ABG73265
ID ABG73265 standard; protein; 305 AA.

AC ABG73265;
DT 17-APR-2003 (first entry)
DE Salmonella typhi cytolsin A (ClyA) protein.
KW Protein expression vector; expression cassette; export protein;
KW transformed host cell; bacterial cell; immune response; HlyE family;
KW cytolsin A; ClyA; recombinant protein; bioreactor; bacterial growth;
KW live vector immunogenic composition; animal feed; animal vaccination;
KW food industry; nutritional supplement; biomediation; waste disposal;
KW waste treatment; host bacterium.

OS Salmonella typhi.
XX US2002146430-A1.
PD 10-OCT-2002.
XX 23-NOV-2001; 2001US-00993292.
XX 22-NOV-2000; 2000US-0252516P.
XX (GALE/) GALEN J E.
PI Galen JE;
XX WPI; 2003-228013/22.
DR N-FSDB; ABX15174.

Novel system for expressing protein of interest, has expression vector with cassette having export protein coding sequence fused to protein coding sequence, host cell transformed with vector, and culturing environment.
PS Claim 5; Page 18; 33pp; English.

The present invention relates to a system for expressing a protein of interest. The system comprises an expression vector having an expression cassette, where the expression cassette comprises an export protein coding sequence genetically fused to a DNA sequence encoding the protein of interest, a host cell transformed with the expression vector, and a culturing environment for the transformed host cell, where the expression cassette expresses the export protein-protein of interest fusion protein which is exported out of the host cell. The system is useful for expressing a gene in a bacterial cell, by providing an expression vector to a population of untransformed bacterial cells, and expressing the expression cassette, such that the export protein-protein of interest fusion protein is produced and exported into the culture medium. The protein of interest is preferably an antigen and this method is useful for eliciting an immune response from a host against the fusion protein. The export protein may be a member of the HlyE family such as cytolsin A (ClyA). The system is useful for the production of recombinant proteins inside a bacterial host cell, in a bioreactor, and various devices that facilitate the growth of bacteria. The system is also useful for providing an animal antigenic material against which an immune response may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can be used as a live vector immunogenic composition capable of facilitating the generation of an immune response in an animal. The system is also useful for preparing immunogenic compositions for vaccinating animals, and for use in the food, nutritional supplement, animal feed, biomediation, waste disposal, and waste treatment industries. The system

is very advantageous for protein expression, as proteins that are toxic to host bacterium can also be expressed. A population of recombinant host cells can be transformed with the expression vector, and then the population of bacterial host cells can be maintained in culture and used to produce proteins without having to harvest and lyse the host cells. The present represents Salmonella typhi ClyA protein

Sequence 305 AA;
Query Match 97.1%; Score 1466; DB 6; Length 305;
Best Local Similarity 97.0%; Pred. No. 7.1e-114;
Matches 294; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 1 MTGIFAEQTVEVVKSAIETADGALDFYKYLQDVI PKTFTDETIKELSRPKQEYSQASV 60
DB 1 MTSIFAEQTVEVVKSAIETADGALDFYKYLQDVI PKTFTDETIKELSRPKQEYSQASV 60
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWGVTQLLSAYILLFDEYNEKKASAKDILIRIL 120
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWGVTQLLSAYILLFDEYNEKKASAKDILIRIL 120
QY 121 DDGVNKLNEAKSLGSSQSFNNASGKLLALDSOLTNDSEKSSYFOSQVDRIKEAYAG 180
DB 121 DDGVNKLNEAKSLGSSQSFNNASGKLLALDSOLTNDSEKSSYFOSQVDRIKEAYAG 180
QY 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQANKDIDA 240
DB 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQANKDIDA 240
QY 241 LKLAETIAAIGEIKTETETTRFFYVDYDLM LSLKGAAKMINTCNEYQQRHGKKTLL 300
DB 241 LKLAETIAAIGEIKTETETTRFFYVDYDLM LSLKGAAKMINTCNEYQQRHGKKTLL 300
QY 301 PDI 303
DB 301 PDV 303

RESULT 2
ABG73266
ID ABG73266 standard; protein; 749 AA.
XX ABG73266;
XX 17-APR-2003 (first entry)
DE Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.

Protein expression vector; expression cassette; export protein; transformed host cell; bacterial cell; immune response; HlyE family; cytolsin A; ClyA; recombinant protein; bioreactor; bacterial growth; live vector immunogenic composition; animal feed; animal vaccination; food industry; nutritional supplement; biomediation; waste disposal; waste treatment; host bacterium; sacB; mutant; mutein; lethal levansucrase.

OS Salmonella typhi.
OS Bacillus subtilis.
OS Synthetic.
OS Chimeric.
PN US2002146430-A1.
PD 10-OCT-2002.
XX 23-NOV-2001; 2001US-00993292.
XX 22-NOV-2000; 2000US-0252516P.
XX (GALE/) GALEN J E.
XX Galen JE;
XX

DR WPI; 2003-228013/22.
DR N-PSDB; ABX15191.

XX Novel system for expressing protein of interest, has expression vector
PT with cassette having export protein coding sequence fused to protein
PT coding sequence, host cell transformed with vector, and culturing
PT environment.

XX Example 3; Page 25-28; 33pp; English.

XX The present invention relates to a system for expressing a protein of
CC interest. The system comprises an expression vector having an expression
CC cassette, where the expression cassette comprises an export protein
CC coding sequence genetically fused to a DNA sequence encoding the protein
CC of interest, a host cell transformed with the expression vector, and a
CC culturing environment for the transformed host cell, where the expression
CC cassette expresses the export protein-protein of interest fusion protein
CC which is exported out of the host cell. The system is useful for
CC expressing a gene in a bacterial cell, by providing an expression vector
CC to a population of untransformed bacterial cells, and expressing the
CC expression cassette, such that the export protein-protein of interest
CC fusion protein is produced and exported into the culture medium. The
CC protein of interest is preferably an antigen and this method is useful
CC for eliciting an immune response from a host against the fusion protein.
CC The export protein may be a member of the HlyE family such as cytolysin A
CC (ClyA). The system is useful for the production of recombinant proteins
CC inside a bacterial host cell, in a bioreactor, and various devices that
CC facilitate the growth of bacteria. The system is also useful for
CC providing an animal antigenic material against which an immune response
CC may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can
CC be used as a live vector immunogenic composition capable of facilitating
CC the generation of an immune response in an animal. The system is also
CC useful for preparing immunogenic compositions for vaccinating animals,
CC and for use in the food, nutritional supplement, animal feed,
CC biomediation, waste disposal, and waste treatment industries. The system
CC is very advantageous for protein expression, as proteins that are toxic
CC to host bacterium can also be expressed. A population of recombinant host
CC cells can be transformed with the expression vector, and then the
CC population of bacterial host cells can be maintained in culture and used
CC to produce proteins without having to harvest and lyse the host cells.
CC The present sequence represents Salmonella typhi ClyA-Bacillus subtilis
CC lethal levanucrase (encoded by sacB gene) fusion protein. The fusion
CC protein is used to investigate the versatility of ClyA as a fusion
CC partner to export heterologous antigens out of S. typhi

XX Sequence 749 AA;

Query Match 97.1%; Score 1466; DB 6; Length 749;
Best Local Similarity 97.0%; Pred. No. 2.3e-113;
Matches 294; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60
DB 1 MTSIFAEQTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILIRIL 120
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILIRIL 120
QY 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAG 180
DB 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAG 180
QY 181 AAAGVAGPFGILLISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAAK 240
DB 181 AAAGVAGPFGILLISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSVTVKQANKDIDAAK 240
QY 241 LKLAIEAIAIGEIKTETETTRFYVDYDDMLSLKGAAKMINTCNEYQQRHGKKTLLLEV 300
DB 241 LKLAIEAIAIGEIKTETETTRFYVDYDDMLSLKGAAKMINTCNEYQQRHGKKTLLLEV 300
QY 301 PDI 303
|||

DB 301 PDV 303

RESULT 3

AAAR67645
ID AAR67645 standard; protein; 309 AA.

XX AC AAR67645;

XX DT 25-MAR-2003 (revised)

XX DT 06-JUL-1995 (first entry)

XX DE Haemolysin regulator.

XX KW Haemolysin regulator; tuberculosis; vaccine; therapy.

XX OS Escherichia coli.

XX PN WO9428137-A1.

XX PD 08-DEC-1994.

XX PF 24-MAY-1994; 94WO-US005869.

XX PR 24-MAY-1993; 93US-00066830.

XX PA (USSH) US DRPT HEALTH & HUMAN SERVICES.

XX PI King CH, Shinnick TM, Sathish M;

XX WPI; 1995-022809/03.

DR P-PSDB; AAQ75857.

XX New nucleic acid encoding haemolytic regulator of E coli - and derived
PT vectors, proteins etc used to transform Mycobacterium cells to produce
PT antituberculosis vaccines providing improved immune response.

XX Claim 9; Page 39-40; 46pp; English.

XX The sequence represents the E. coli haemolysis regulator which may be
CC used as an immunogen in a vaccine composition against tuberculosis
CC (Mycobacterium tuberculosis). (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-
CC 2003 to correct PA field.)

XX SQ Sequence 309 AA;

Query Match 90.7%; Score 1370; DB 2; Length 309;
Best Local Similarity 90.0%; Pred. No. 7.3e-106;
Matches 271; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MTGIFAEQTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60
DB 1 MTEIVADKTVVVKSAIETADGALDFYKNYLDQVTPWKTFTDTIKELSRFKQEQYSQASV 60
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILIRIL 120
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILIRIL 120
QY 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAG 180
DB 121 DDGVNKLNEAOKSLILGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKEAYAG 180
QY 181 AAAGVAGPFGILLISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAAK 240
DB 181 AAAGVAGPFGILLISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSVTVKQANKDIDAAK 240
QY 241 LKLAIEAIAIGEIKTETETTRFYVDYDDMLSLKGAAKMINTCNEYQQRHGKKTLLLEV 300
DB 241 LKLAIEAIAIGEIKTETETTRFYVDYDDMLSLKGAAKMINTCNEYQQRHGKKTLLLEV 300
QY 301 P 301
|||

Db 301 P 301

RESULT 4
ABG25221
ID ABG25221 standard; protein; 1035 AA.
XX
AC ABG25221;
XX
DT 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #25212.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
KW
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US008631.
PF
XX 31-MAR-2000; 2000US-00540217.
PR
XX 23-AUG-2000; 2000US-00649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS89408.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 55580; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities in
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1035 AA;
SQ

Query Match 70.5%; Score 1065; DB 4; Length 1035;
Best Local Similarity 91.3%; Pred. No. 9.8e-80;
Matches 210; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 70 MDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKQKILIRLDDGVNKLNE 129
DB 1 MDSQDKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKASAKQKILIRLDDGVNKLNE 60

QY 130 AQKSLIGSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDRIKKEAYAGAAGIVAGP 189
DB 61 AQKSLIVSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAGVAGP 120
QY 190 FGLIISYSIAAGVIECKLIPELNDRKAVONFETSLVTVKQANKDIDAOKLKLATEIAA 249
DB 121 FGLIISYSIAAGVIECKLIPELNDRKAVONFETSLVTVKQANKDIDAOKLKLATEIAA 180
QY 250 IGEIKTETETTFYVDYDDMLSLKGAARKMINTCNEYQQRHGKKTLE 299
DB 181 IGEIKTETETTFYVDYDDMLSLKGAARKMINTCNEYQQRHGKKTLE 230

RESULT 5
ABM68103
ID ABM68103 standard; protein; 582 AA.
XX
AC ABM68103;
XX
DT 20-NOV-2003 (first entry)
XX
XX Photorhabdus luminescens protein sequence #1200.
DE
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
XX Photorhabdus luminescens.
OS
XX WO200294867-A2.
FN
XX 28-NOV-2002.
PD
XX 07-FEB-2002; 2002WO-IB003040.
PF
XX 07-FEB-2001; 2001FR-00001659.
PR
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
PA
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
PI
XX WPI; 2003-148459/14.
DR
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
PT
XX Claim 2; SEQ ID NO 1200; 1205pp; French.
PS
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX

SQ Sequence 582 AA;
Query Match 8.4%; Score 126.5; DB 6; Length 582;
Best Local Similarity 19.0%; Pred. No. 0.089;
Matches 80; Conservative 55; Mismatches 132; Indels 155; Gaps 14;
QY 1 MTGIPAEQTVVVK-----SAIETAGALDFYKYLDOV-----IPWKTF----- 40
DB 107 ITGDFSKGFDLVNFKMLPNRGDGNLETPPPAYFLPFYDQIKSSWSPNSENLCQY 166
QY 41 -----DETIELSRFKOEYSOEASVLVCDIKVL----- 68
DB 167 SNWKGPLIKYFTGYLKPEHFDIENIYSEIKKSAHKIEKFOSAVEVIVDNTVDSPT 226
QY 69 -----LMSDQKYPEATQTVYEWGVVTVQLLSAYILLFDEYNKKA 109
DB 227 LDNNDFFIKQNEIKNELYLDIDYQRLVDAQTI---TSNIYDLESQYALAITSALEA 283
QY 110 -----SAQKDL-----TRILDDGVNKLNEAKSKLLGSSQSFNNASG 146
DB 284 DYKFAVESIPTDYLECPCLGTHDNLNPNRALLSEKDSLNEA-----NSTAS 332
QY 147 KLLALDSQLTWDFSEKSYFOSQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGK 206
DB 333 KIVELKSSL-NSLNEDAQFTNEIERINN-----YITDDDAVKKG 372
QY 207 LIPELNDRL-----KAVQNFTSLSVTVKQANKDI-----DAAKL----- 241
DB 373 LIAQVIDTLSTENVSKNIQIKIDNEDLNISKANSIKELKDKQKLLSTKEKEELNSSFM 432
QY 242 -KLATEIAIGEIKTETETTPFYVDYDMLSLKGAAKMI-NTCNFYQORHGKKTILE 299
DB 433 SKLLANIEALGSTGINLSKVSPTDYKQLGGGAEEAARGELAYQLSVLQIIGHAKTCV 492
QY 300 VP 301
DB 493 PP 494
RESULT 6
ID ABU18419 standard; protein; 941 AA.
AC ABU18419;
DT 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #3946.
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Bacillus anthracis.
OS WO200277183-A2.
FN 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA22289.
XX

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
PS Claim 25; SEQ ID NO 46343; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

Sequence 941 AA;

Query Match 7.7%; Score 116; DB 6; Length 941;
Best Local Similarity 20.1%; Pred. No. 1.3;
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7;
QY 7 EQTVVVKSAIETADGAL-DFYKYLDOVIP-WKTFDETIKELSRFKOEYSOEASVLVGD 64
DB 596 KEVVESEINEATQNASAQLGDFLATYDSEIVNFNTAERTKMSKNTSAILKEADKKLPD 655
QY 65 IKVLLMDSQDKYFEATQTVYEWGVVTVQLLSAYILLFDEYNKASAKQDLIRLLDDGV 124
DB 656 VKLLEDSK-----GLV-----DGR 671
QY 125 NKLNEAKSKLLGSSQSFNNASGKLLALDSQ-----LTNDFSEKSSYFOSQVDRIRK 175
DB 672 KGLADIKAEMPAETKKIKELADKIRDFESEEDLKDIIRLLKNDVKEQSDYFANPNLKEN 731
QY 176 EAYAGAAGIVAGPGLIISYSIAAGVIEGKLIPELND---RLKAVQNFP-----TSLSVT 228
DB 732 KLFAMPNYSAMSFPYTVIALWVALLVSVLLTVVEHEGANYKSHIEYFGRLTLFTLTMG 791
QY 229 VKQA 232
DB 792 LSQA 795
RESULT 7
ID ABP76825 standard; protein; 1015 AA.
XX AC ABP76825;
XX DT 07-MAR-2003 (first entry)
XX

DE N. gonorrhoeae amino acid sequence SEQ ID 180.
XX Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX 12-FEB-2002; 2002WO-IB002069.
XX 12-FEB-2001; 2001GB-00003424.
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ37795.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX Claim 1; Page 197; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention
XX
XX Sequence 1015 AA;

Query Match 7.5%; Score 113; DB 6; Length 1015;
Best Local Similarity 21.8%; Pred. No. 2.5;
Matches 75; Conservative 54; Mismatches 113; Indels 102; Gaps 17;
QY 9 TVEVVKSAIETADGALDFYNYKLDQVTPWKT-----FDETIKELSRFKQBYSQEASVLVG 63
Db 104 TADFAKSLVLTAD-AMQINSQVRQVTSSETEYLAVQQQLDNTANRTRASLESTANLYVS 162
QY 64 DIKVL--MDSQD--KYFEATQTVYEWCGV-----VTOLLGAYILLDFEYNEK 107
Db 163 TGRALKDYGVTQOEILKFTFAANNAMTIGVGAAQQAALMQLSQAALGSGVLQGEF--K 220
QY 108 KASAKDILIRIL-----DDGVNKL-NEAQ-----KSLGSSQSFNNASGKLALD 152
Db 221 SISEAAPILLDTIAEYMGKSRDEIKLGSEGLTADVIFKAISGASEKFGQEAARMPVTM 280
QY 153 SQTNDPSEKSSYFQSOVDRIKE-----AYAGAAA- 183
Db 281 GOALTTFVSN--WQSVKSLNDSGTMSGIAAVIKLIADNLNLVPIVAGFAVAVA 337
QY 184 -----GIVAGPGLIISYTAAGVIRGKLIPELNDRLKAVQNFSTLSVTVKQAKND 235
Db 338 APTIALNLALLANPFGII---AVAIGTVVG-LIAKFGDEIDVFGGWSNLS-----D 385
QY 236 IDAAKLKLAIEA--AIGEIKTETETTFYVDYDDLMLSLKGA 277
Db 386 VIRAVQIITETVGEAVGTVKS-----PDGLTGRINEGA 420

RESULT 8
ABP78617
ID ABP78617 standard; protein; 1015 AA.
XX
XX AC ABP78617;
XX DT 07-MAR-2003 (first entry)

XX N. gonorrhoeae amino acid sequence SEQ ID 3764.
XX Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX 12-FEB-2002; 2002WO-IB002069.
XX 12-FEB-2001; 2001GB-00003424.
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ39587.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX Disclosure; Page 460; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX
XX Sequence 1015 AA;

Query Match 7.5%; Score 113; DB 6; Length 1015;
Best Local Similarity 21.8%; Pred. No. 2.5;
Matches 75; Conservative 54; Mismatches 113; Indels 102; Gaps 17;
QY 9 TVEVVKSAIETADGALDFYNYKLDQVTPWKT-----FDETIKELSRFKQBYSQEASVLVG 63
Db 104 TADFAKSLVLTAD-AMQINSQVRQVTSSETEYLAVQQQLDNTANRTRASLESTANLYVS 162
QY 64 DIKVL--MDSQD--KYFEATQTVYEWCGV-----VTOLLGAYILLDFEYNEK 107
Db 163 TGRALKDYGVTQOEILKFTFAANNAMTIGVGAAQQAALMQLSQAALGSGVLQGEF--K 220
QY 108 KASAKDILIRIL-----DDGVNKL-NEAQ-----KSLGSSQSFNNASGKLALD 152
Db 221 SISEAAPILLDTIAEYMGKSRDEIKLGSEGLTADVIFKAISGASEKFGQEAARMPVTM 280
QY 153 SQTNDPSEKSSYFQSOVDRIKE-----AYAGAAA- 183
Db 281 GOALTTFVSN--WQSVKSLNDSGTMSGIAAVIKLIADNLNLVPIVAGFAVAVA 337
QY 184 -----GIVAGPGLIISYTAAGVIRGKLIPELNDRLKAVQNFSTLSVTVKQAKND 235
Db 338 APTIALNLALLANPFGII---AVAIGTVVG-LIAKFGDEIDVFGGWSNLS-----D 385
QY 236 IDAAKLKLAIEA--AIGEIKTETETTFYVDYDDLMLSLKGA 277
Db 386 VIRAVQIITETVGEAVGTVKS-----PDGLTGRINEGA 420

RESULT 9
ADF07451
ID ADF07451 standard; protein; 1036 AA.
XX
XX AC ADF07451;
XX

DT 12-FEB-2004 (first entry)
XX Bacterial polypeptide #3564.
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant.
XX
XX Proteus mirabilis.
OS
PN US6605709-B1.
XX
XX 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
PF
XX
XX 09-APR-1999; 99US-0128706P.
PR
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
PI
XX
XX WPI; 2003-895291/82.
DR N-PSDB; ADF03279.
XX

XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 7736; 870pp; English.
PS
XX
XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polynucleotides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis, a
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
XX Sequence 1036 AA;

Query Match 7.4%; Score 111.5; DB 7; Length 1036;
Best Local Similarity 21.0%; Pred. No. 3.4;
Matches 69; Conservative 62; Mismatches 117; Indels 81; Gaps 14;

Qy 2 TGIFAEQTVVVK-----SALTADGALDFVNYKLDQVLPWKTFDETIKELSRKQE 53
Db 167 TALSGMEINELIKKQNGENVSSSELAKASIDLNLQVD-----TASLNNNISAFSQ 220
Qy 54 YSQAESVL-----VGDIKVLLMDSQDKYFEATQTVYEWGCVVVTQLLSAYILL-PDEY 104
Db 221 LNKGLSVLNTKHLNGVGN-KLQNLPLNDKLGTLGDTV---SGILSALSASFILSNADAD 276
Qy 105 NEKASAKDILIRLDD---GVNKLNEAQSLLGSSOSFNNASGLIA-----LD 152
Db 277 TGTAAAGVLTFTKVLGNVGNKAVSYILQRAVQGLSTS--AASAGLIASAVTLAISPLS 334
Qy 153 SOLTNDPSEKSYFQSDVRIRKEAYACAAAGIVAGPGLIISYSIAAGVIEGKLIPELN 212
Db 335 FLATADQPKRANKIEYSQRFKKFYEGDS-----LIAAFKFTGALDASL----- 380
Qy 213 DRLKAVQNFFTLSVTVKQANKDIDAALK--LATEIAAI-----GEIKTETETTRFYVD 265
Db 381 -----TTINTALGTISAGISAASATSLIGAPISALVCAITGILSIGILEASK----- 426
Qy 266 YDDLMLSLKGAQKQINTCNEYQORHOK 294
Db 427 -----QSMFEHVRNMTAIEWEKTHGK 450

RESULT 10
ABJ19025

ID ABJ19025 standard; protein; 2681 AA.

XX AC ABJ19025;

XX DT 06-MAR-2003 (first entry)

XX Pathogen specific antigen related staphylococcal protein SEQ ID No 274.
DE
XX Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
KW autoimmune disease; HIV; hepatitis.
XX

OS Staphylococcus sp.

XX WO200259148-A2;

XX PN 01-AUG-2002.

XX PD 21-JAN-2002; 2002WO-EP000546.

XX PF 26-JAN-2001; 2001AT-00000130.

XX PR (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

XX PA Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;
PI Minh DB, Vytvytska O, Etz H, Dryia A, Weichhart T, Hafner M;
PI Tempelmaier B;
XX
XX WPI; 2003-075410/07.

XX Identifying, isolating and producing hyperimmune serum-reactive antigens
PT from a pathogen, for preparing vaccine or medicament for treating or
PT preventing e.g. staphylococcal infections, comprises providing antibody
PT preparation.

XX Example 7; Page 200; 252pp; English.

XX The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens are
CC used in a vaccine, comprises providing antibody preparation from a plasma
CC pool of a type of animal, or individual sera with antibodies against the
CC specific pathogen, tumour, allergen, tissue or host prone to auto-
CC immunity. The hyperimmune serum-reactive antigens comprising any of the
CC 62 sequences of 53-2261 amino acids fully defined in the specification,
CC or their hyperimmune fragments are useful for the manufacture of a
CC pharmaceutical preparation, particularly a vaccine against staphylococcal
CC infections or colonisation against S. aureus or S. epidermidis. The
CC preparation of antibodies is useful for the manufacture of a medicament
CC for treating or preventing staphylococcal infections or colonisation
CC against S. aureus or S. epidermidis. The antibody preparations may also
CC be used for diagnostic and imaging purposes. Other conditions that can be
CC treated include cancer, autoimmune diseases or infections caused by viral
CC (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This
CC sequence represents a staphylococcal protein relating to the method for
CC identifying and producing pathogen specific antigens of the invention
XX

XX Sequence 2681 AA;

Query Match 7.3%; Score 110.5; DB 6; Length 2681;

Best Local Similarity 20.7%; Pred. No. 14;

Matches 71; Conservative 52; Mismatches 129; Indels 91; Gaps 13;

Qy 2 TGIFAEQTVVVKSAIETADGALDFVNYKLDQVLPWKTFDETIKELSR 49
Db 998 TDLKSEKLDLTKS-YKLTDTSFNDVFKYIDEMTDRIADTEKVNKKYKDYTLQKNLSD 1056

Qy 50 FKQBY-----SQEASVLVGDIKVL-----LMSDQDKYFEATQTVYEWG--G 88

Db 1057 YMKYDNCILEISKYSNDAADKVLGDTATATELQNDPQVQKONWAEFKQITLESFQDG 1116
 Qy 89 VTQLLSAYILLDFDYNEKKASAKDILIRLLDDGVNKNLNEAQAQKSLGSSQSFNNASGL 148
 Db 1117 IVT-----BAEARLRVQLDMLDRESMD-----IEERYKSLLANQVTTNDIKNRL 1161
 Qy 149 LALDSQLTDFSEKSSYFSQSDVRIRKEAYAGAAAGIVAGPGLIISYSIAAGVLEGKLI 208
 Db 1162 TA-----SRPFYLSVHAS-LRK-----VIEQIIADGRVDESEK 1193
 Qy 209 PELNDRKAVONFFTSLSVTYKQANKOI-----DAAKLKLATEIAAIGETETETTRF 262
 Db 1194 TLANNSLNTYNTTLAYSKTIQEAQNTLSQIISDVASKKVEEFGVITTISSDVTDTIKK 1253
 Qy 263 YVD-----YDDLMLSLKGAKKMINTCNEYQQRHGKKTLL 299
 Db 1254 QRDGAVITYYSGVPTLSNDPAKSW--TTNDLKOLHIKDMYLD 1294

RESULT 11

ABU42996

ID ABU42996 standard; protein; 1189 AA.

AC ABU42996;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #28523.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Staphylococcus epidermidis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACN46866.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 70920; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation or the activity of a gene in an operon required for

XX proliferation; (7) identifying a compound that influences the activity of

XX the gene product or that has an activity against a biological pathway

XX required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 1189 AA;

Query Match

Best Local Similarity 7.3%; Score 110; DB 6; Length 1189;

Matches 61; Conservative 18.3%; Pred. No. 5.4; Mismatches 60; Indels 82; Gaps 9;

Qy 7 EQTVEVVK-----SAIETADGALDFYKYLQDVIPWKTDETIKELSRFKQYSQASV 60

Db 163 EESAGVLKYKKAESIQKLDHTEDNLNRVEDILYDLGRVPELKEAAIAKEYKQLSKE 222

Qy 61 LVGDIKVLMDSQKYFEATQTVYEMC-----GVVTOLLSAVILLDFEYNE 106

Db 223 MEQSDVITVSDIDHYTENDQRLDERLNHLKSOQAQKEGQQAQINQLQRY----- 273

Qy 107 KASAAQDILIRLLDGVNKLNEAQAQKSLGSSQSFNNASGKLLALDLSQLTND-----FS 160

Db 274 -GKRQON-----DYDIEKLN-----YELVATENEYQSGKLVLEERKQKQSTNARYE 323

Qy 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNDRL-----K 216

Db 324 EELDNLESQIDSINKKAQNE-----KLLADLKNQKQKQLNK 359

Qy 217 AVQNFSTLSVTYVQANKDDAAKLKLAETAAIGETETETTRFYVDYDMLSLKLG 276

Db 360 EVQELSLLYVSDQHDQKLEIKNSYITLMSQSDVNNDIRFLEHTINENEAKKSRDLS 419

Qy 277 AAKMINTCNEYQO-----RHGKKTLL 300

Db 420 RLVEAFNLQDKIQNITQTNKAYOSSKSKMOKV 452

RESULT 12

ADC95690

ID ADC95690 standard; protein; 515 AA.

XX AC ADC95690;

XX 01-JAN-2004 (first entry)

XX E. faecium protein sequence SEQ ID 5317.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;

XX abdominal-pelvic infection.

XX Enterococcus faecium.

XX US583275-B1.

XX 24-JUN-2003.

XX 30-JUN-1998; 98US-00107532.

XX 02-JUL-1997; 97US-0051571P.

XX 14-MAY-1998; 98US-0085598P.

Db 325 QTSIAEAKQ-----KVK-----HPEKLOESLMKAAAEKETIQ 357

Qy 281 -----WINTCNEYQ--QRHGKKTLLLEVDP 302

Db 358 KAEANLIKTQOELEKYQKSTKELLALRD 386

RESULT 14

ABU25599

ID ABU25599 standard; protein; 742 AA.

XX AC ABU25599;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #11126.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX KW Clostridium difficile.

OS WO200277183-A2.

XX PN 03-OCT-2002.

XX PD 21-MAR-2002; 2002WO-US009107.

XX PF 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR N-PSDB; ACA29469.

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA29469.

XX PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 53523; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 742 AA;

Query Match 7.1%; Score 107.5; DB 6; Length 742;

Best Local Similarity 21.2%; Pred. No. 4.7; Mismatches 63; Gaps 10;

Matches 66; Conservative 49; Indels 134; Indels 63; Gaps 10;

Qy 18 ETADGALDFYKYLDOVIPKTFDETETIKELSRFQEQSVLVDIKVLMDSQDKYF 77

Db 236 KTVDLASDATSKLSDTVKDIKSDLPITKKT-----LNDTKLLSSDLKKFLEDNDNDL 288

Qy 78 BATQTVVWCGVTVQLLSAYILLFDEYNEKKAQKQILIRLD-----DGVNKLNEAQ 131

Db 289 ELSPLIKSDLMVLDL-----SSASSLTLLIDAVNSGSDVPKLDNL 333

Qy 132 KSLIGSSQSFNNASGKLLALDSQLTN-----DFSEKSSYFQSDVDRIRKEAYA 179

Db 334 SEKLSNLQSLNDTLVDFLTQKLNQLTNNRLDDVDINDLEDSSNKIDSSISTINDIKNKVIS 393

Qy 180 GAAAGIVAGPFGLIISYIAAGVIEGKLIPELNDRL-KAVQNFFTSLSVTVQAKMDI-- 236

Db 394 GQPSISA--LNNVLSLSNGIGRINLNLNFDKSKIPINNIFAN-----SIKVANDIITV 448

Qy 237 -DAAKLK-----LATEIAAIGEIKTETETRTFYVDYDDMLSLKGAAKKMINTCNEY 288

Db 449 LDKAEAKLPKVEEILTTSKLKSGNAQBSISLIR-----BRLPLAKGMLDLDLITLSKI 501

Qy 289 QQRHGKKTLLLEV 300

Db 502 SNGEDMKKLVSL 513

RESULT 15

ABU43311

ID ABU43311 standard; protein; 3692 AA.

XX AC ABU43311;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #28838.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Staphylococcus epidermidis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA47181.

XX PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 71235; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 3692 AA;

Query Match 7.1%; Score 106.5; DB 6; Length 3692;

Best Local Similarity 21.5%; Pred. No. 46;

Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;

Qy 11 EVKSAIETADGALDFYKYLDDVPIPKTPTDETIKELSR-----FKQYSQ-EA 58
Db 564 EQVNDIIPSNYTLASYNKY-----NKLKERAQTVLDBETNTPNQRYSQTQI 612
Qy 59 SVLVGDIKVLIM-----DSQDKYFATQTVYEWCGVVTQLLSAYILLFPEYNEKKASA 111
Db 613 DLLLHELOTTLINRVASREINDKAQEMTDAVYDSTELTTEKDT---LVQDIENHNEI 669
Qy 112 QKILIRILDGVNKLNEAQSLLGSS-----QSFNNSAGKLLALDSQLTNDFS 160
Db 670 SNNIDDELTDGVERVKEAGLHTLESPTPHVTPKPNARQVNNRA-----DQCKTLIRN 723
Qy 161 EKSSYFOSQVDRIRK-EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQ 219
Db 724 NHEATTEQNEAIRQVEAHSDDA-----IAKIGEAETDTTNE---ARD 764
Qy 220 NFFTSLSVTVKQANKDIDAAKILKLAETAAIGETETETTRFYVDVDDMLSLKGAAK 279
Db 765 NGTKLIATDVNPPTKAAE-----RAAVTNSANSKIKDINNNTQATLDERDAIALVNRSD 821
Qy 280 KMINTCNEYQ-----QRHGKKTLLFVP 301
Db 822 EAIQNINTAQGNDVDVTEAQNGTNTIQQVP 851

RESULT 16

ABP40235

XX ABP40235 standard; protein; 3696 AA.

AC ABP40235;

XX

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.

KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.

OS Staphylococcus epidermidis.

PN US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.

XX 14-AUG-1997; 97US-0055779P.

PR 08-NOV-1997; 97US-0064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2002-381255/41.

XX N-PSDB; ABN92780.

PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.

XX Disclosure; SEQ ID NO 5080; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly *S. epidermidis* infections. The sequences can be used to
CC screen for compounds able to interfere with the *S. epidermidis* life cycle
CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site

XX SQ Sequence 3696 AA;

Query Match 7.1%; Score 106.5; DB 5; Length 3696;

Best Local Similarity 21.5%; Pred. No. 47;

Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;

Qy 11 EVKSAIETADGALDFYKYLDDVPIPKTPTDETIKELSR-----FKQYSQ-EA 58
Db 568 EQVNDIIPSNYTLASYNKY-----NKLKERAQTVLDBETNTPNQRYSQTQI 616
Qy 59 SVLVGDIKVLIM-----DSQDKYFATQTVYEWCGVVTQLLSAYILLFPEYNEKKASA 111
Db 617 DLLLHELOTTLINRVASREINDKAQEMTDAVYDSTELTTEKDT---LVQDIENHNEI 673
Qy 112 QKILIRILDGVNKLNEAQSLLGSS-----QSFNNSAGKLLALDSQLTNDFS 160
Db 674 SNNIDDELTDGVERVKEAGLHTLESPTPHVTPKPNARQVNNRA-----DQCKTLIRN 727
Qy 161 EKSSYFOSQVDRIRK-EAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNDRLKAVQ 219
Db 728 NHEATTEQNEAIRQVEAHSDDA-----IAKIGEAETDTTNE---ARD 768
Qy 220 NFFTSLSVTVKQANKDIDAAKILKLAETAAIGETETETTRFYVDVDDMLSLKGAAK 279
Db 769 NGTKLIATDVNPPTKAAE-----RAAVTNSANSKIKDINNNTQATLDERDAIALVNRSD 825
Qy 280 KMINTCNEYQ-----QRHGKKTLLFVP 301
Db 826 EAIQNINTAQGNDVDVTEAQNGTNTIQQVP 855

RESULT 17

AD95523
 ID ADC95523 standard; protein; 1231 AA.
 XX AC ADC95523;
 XX DT 01-JAN-2004 (first entry)
 XX DE E. faecium protein sequence SEQ ID 5150.
 XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 XX KW abdominal-pelvic infection.
 XX OS Enterococcus faecium.
 XX PN US5583275-B1.
 XX PD 24-JUN-2003.
 XX PF 30-JUN-1998; 98US-00107532.
 XX PR 02-JUL-1997; 97US-0051571P.
 XX PR 14-MAY-1998; 98US-0085598P.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 XX PA Doucette-Stamm LA, Bush D;
 XX PI WPI; 2003-799836/75.
 XX DR N-PSDB; ADC91869.
 XX New isolated nucleic acid derived from Enterococcus faecium encoding an
 PT Enterococcus faecium polypeptide useful for detection, prevention and
 PT treatment of a pathological condition resulting from a bacterial
 PT infection.
 XX Example 1; SEQ ID NO 5150; 243pp; English.
 XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids are useful for diagnosing pathological conditions
 CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of Candida albicans -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.
 XX Sequence 1231 AA;
 SQ
 Query Match 7.0%; Score 106; DB 7; Length 1231;
 Best Local Similarity 23.0%; Pred. No. 12;
 Matches 77; Conservative 60; Mismatches 124; Indels 74; Gaps 17;
 6 ABOYVEVKSATETADGALD-----FKNYLDQVTPKTFDE-----TIKELS 48
 269 ABEIVSMGSVAQTETQTDYLNANGRTGFLNIHLRFPFVTFQKLPHTVKTIAVL 328
 49 RFKQYSQBSAVLGVGDKVLLMDSQ-----DKYFEATQTYVEGCVVQLLSAYILLF 101
 329 RSKEPGAGGEPILL-DVQSAVYSELPAVIGRYGLGSK-----DVTPTDQISA---VF 378
 102 DEYNKKAQAOKDILLRIIDGWNKLNKAQKSL-LGSSQSEN-----NASGKL 148
 379 DELKDKPSIRKKRFTIGIVDDVTYQSLEPKESLDLTPQTFOAKFWGSGDGTGANKSA 438

149 LALDSQLTNDSEKSSYFOSQ-----VDRIR-----KEAYAGAAAGIVA--GPFGLII 194
 439 IKIIGDHTDKYAQGYFYDYSKSGGLTVSHLFGDTPIRSAYLVEHADLVACHTP-AYLH 497
 195 SYSIAAGVIEGKLIPELNDRLKAVQNFSTLSVTVKO--ANKDI---DAAKLKLAATEIAA 249
 498 SYDLVKGKPGGIF--LLNTLWSDEQLETHLPLKLYLAENNIRFYTNAMRLAQEVL 555
 250 IGEIKTETETTRP-----YVDYDDLMLSLKGAANK 280
 556 GRRINTAMETAFFKLADIIPFDE-VLPLKBEALK 589

RESULT 18
 ADE56031
 ID ADE56031 standard; protein; 3187 AA.
 XX AC ADE56031;
 XX DT 29-JAN-2004 (first entry)
 XX DE Rat Protein BAA05026, SEQ ID NO 1870.
 XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 XX KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX OS Rattus norvegicus.
 XX PN WO2003016475-A2.
 XX PD 27-FEB-2003.
 XX PF 14-AUG-2002; 2002WO-US025765.
 XX PR 14-AUG-2001; 2001US-0312147P.
 XX PR 01-NOV-2001; 2001US-0346382P.
 XX PR 26-NOV-2001; 2001US-0333347P.
 XX (GENO) GEN HOSPITAL CORP.
 XX (PARR) BAYER AG.
 XX Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 XX GENBANK; BAA05026.
 XX New composition comprising two or more isolated polypeptides, useful for
 XX preparing a medicament for treating pain in an animal.
 XX Claim 1; Page; 1017pp; English.
 XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

```
CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
```

XX SQ Sequence 3187 AA;

	Query Match	7.0%; Score 106; DB 7; Length 3187;
	Best Local Similarity	22.2%; Pred. No. 42;
	Matches	Conservative 53; Mismatches 134; Indels 86; Gaps 16;

Qy	6	AQTV--EVVKSAIETADGALDFNKKYLDQVIPWKTDETIKELSRFKQYSQBSVLVG	63
Db	1748	AEATLLANSAPGVSETFSFHHDDINNYLQQDLK--GRIALEMEKQKRELSOTLEN	1804
		:	:
Qy	64	DIKVLL-----MDSQKYFE-----ATQTVVGCVVVTOLLNAVILLPFDVEYNKKAS	110
Db	1805	EKNALLTQISAKUSELKLEEVAKINMLNQIQEELSRTVKL-----KETAE	1852
		: : :	:
Qy	111	AQKDILIRILDGGVNKLNEAQSLLGSQSFNNASGKLALDLSOLTN-----DFSSESS	164
Db	1853	EEKODLEERL--MNQLAELNGSIGNYQVDVTAQIKNEQLESEMQLKRCVSELBECK	1909
		: : :	:
Qy	165	YP---QSQVD-RIRKEAYAGAAIGVAPGLIIYSISIAAGVTIEGKLIPELNRLKAQV-	219
Db	1910	QLVKERTKIVSERIKE-YMEKIQAQGPQSGKHAKELQ-----ELLKEQOEVEQLQK	1962
		:::::	:
Qy	220	-----NFFTSLSVTVK-----OANKDIDAACKKLATEI-----AAIGEIK----	254
Db	1963	DCIRYLGRISALETVKVALFHVHTESQKDLDTAKGNLAQAVEHKKQAQELSFKILLDD	2022
		:: :	:
Qy	255	TETETTFRYVD-----YDDLMLSLLGAAKGMINTCNEYQQRHGKK	295
Db	2023	TOSEAAARVLADNLKKLELOSNTKESIKSQIKOKEDLLRLELOAEKHKE	2073
		: : :	:

RESULT 19	
ADE56035	
ID	ADE56035 standard; protein; 3187 AA.
XX	
AC	ADE56035;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Rat Protein BAA05026, SEQ ID NO 1874.
XX	
KW	Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX	chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX	
OS	Rattus norvegicus.
XX	
PN	WO2003016475-A2.
XX	
PD	27-FEB-2003.
PF	
PP	14-AUG-2002; 2002WO-US025765.
XX	
PR	14-AUG-2001; 2001US-0312147P.
PR	01-NOV-2001; 2001US-0346382P.
PR	26-NOV-2001; 2001US-0333347P.
XX	
PA	(GEO) GEN HOSPITAL CORP.
PA	(FARB) BAYER AG.
XX	
PI	Woolf C, D'urso D, Befort K, Costigan M;
XX	
DR	WPI; 2003-268312/26.
DR	GENBANK; BAA05026.
XX	
PT	New composition comprising two or more isolated polypeptides, useful for
PT	preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a rat protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Query Match
Best Local Similarity 7.0%; Score 106; DB 7; Length 3187;
Matches 78; Conservative 53; Mismatches 134; Indels 86; Gaps 16;

QY 6 AEQTV---EVVKSAIETADGALDPFNKYLDQVIPWKTPTDETIKLSRFKQBSQEASVLVG 63
DB 1748 AEATLLANGAKPGVSETPSSHDDINNYLQDLQLK--GRITAELEMEKQXDRSLSQTLN 1804

QY 64 DIKVLL-----MDSQKYFE-----ATQTVYEWGVVTOLLISAYILLFDEYNKKAS 110
DB 1805 EKNAALLTQISAQSDSELKLEEEVYAKIMNLNQIQEELSRTKL-----KETAE 1852

QY 111 AQKDILIRILDGCVKNLEAQSGLLGSSQFNAGSKLLDALDLSOLTN-----DFSEKSS 164
DB 1853 EEKODLEERU---MNQLAELNGSIGNYQVDTPAQIKNEGLESEMQLKRCVSELEEKG 1909

QY 165 YF---QSQVD-RIRKEAYAGAAAAGIVAGPGLIIYSIAAGVIEGKLIPELNRLKAVQ- 219
DB 1910 QLVKENTKVESEIRKE-YMEKIQAQKGPCKSHAKELQ-----ELLKEQQEBVKQLQK 1962

QY 220 -----NFTSLSVTVK-----QANKDIDAAKLKLATETI-----AAIGEK----- 254
DB 1963 DCIRYLGRISALEKTVKALEFVFHTTESQOKLDATKGNLAQAVEHHKKAQAELSFPKILLDD 2022

QY 255 TETETTRFYVD-----YDDLMLSLLKGAACKMINTCNYQVRHGKK 295
DB 2023 TQSEAARVLADNLKKELOSNEKESIKSIQKQDEDLLRLEQAEKHRE 2073

RESULT 20
ABB71280
ID ABB71280 standard; protein; 588 AA.
XX ABB71280;
XX AC
XX DT
XX DE
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40632.
XX XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL15383.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Disclosure; SEQ ID NO 40632; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 588 AA;
XX
Query Match 7.0%; Score 105.5; DB 4; Length 588;
Best Local Similarity 20.5%; Pred. No. 5.1;
Matches 66; Conservative 44; Mismatches 87; Indels 125; Gaps 13;
OY 11 EVVKGAIETADGALDFYNYLDQV--IPWK----- 38
DB 300 EIIALLAENAGALDRFNTQLDVAAGINWKIDAEIDNQAELYDAITEVTEELLRNQTELL 359
OY 39 -----TFD-----ETIKELSRFKQEVQSEASVLYGDIKULLMDSQDKY 76
DB 360 LWEVELLRGVVTSFQNLDFANRSIEAVSDTLRLOEQNKDRVRLNLDKLVN----TQGOI 415
OY 77 FEATQTVYEWCGVVTOLLISAYILLPDEYNEKKASAKQDILIRLDDGVNKLNEAQ----- 131
DB 416 LRRTKGLDRLNLFVQLLGYI-----EPKNS-----LSDSFDNLKNSQINSLI 460
OY 132 --KSLIGSSQSFNNASGKLLALDLSQITNDFSEKSSYFQSDVRIRKEAYAGAAAGIVAGP 189
DB 461 ELKNVPEVRNLTKTSIRKLSFLDNQL-----ALFNQIQENRY----- 497
OY 190 FGLIISYSTAAGVIEGKLIPELNDRIKAVQNFSTLSVTVKQANKDIDAKLIKATETAA 249
DB 498 -----YSVEAVI---KAWTPTN--LKEINDLTHALSISQKRTD-----LAIAISG 537
OY 250 IGEIKTETETTRF--VVDYDDL 269
DB 538 SAEYNTETPTPTFISYKGIEDI 559
RESULT 21
ID ABB77986
XX ABB77986 standard; protein; 1292 AA.
XX AC ABB77986;

XX 22-OCT-2002 (first entry)
XX Amino acid sequence of an Arabidopsis RAD50 homologue.
XX Nucleic acid integration; homologous recombination; telomeric region;
XX RAD50.
XX Arabidopsis thaliana.
XX EPI217074-A1.
XX 26-JUN-2002.
XX 22-DEC-2000; 2000EP-00204693.
XX 22-DEC-2000; 2000EP-00204693.
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (BINA-) STICHTING BINAIR VECTOR SYSTEEM.
XX Hooykaas PJJ, Van Attikum H, Bundock P;
XX WPI; 2002-550409/59.
XX Directing integration of nucleic acid of interest to a sub-telomeric
XX region in an eukaryote with preference for non-homologous recombination,
XX by steering an integration pathway towards homologous recombination.
XX Disclosure; Fig 5; 63pp; English.
XX The specification describes a method for directing integration of a
XX nucleic acid of interest to a pre-determined site, where the nucleic acid
XX has homology at or around the pre-determined site, in a eukaryote with a
XX preference for non-homologous recombination. The method comprises
XX steering an integration pathway towards homologous recombination. The
XX method is useful for directing integration of a nucleic acid of interest
XX to a subtelomeric and/or telomeric region in an eukaryote with a
XX preference for non-homologous recombination. The nucleic acid of interest
XX comprises an inactive gene to replace an active gene, or vice versa, is a
XX portion of a gene delivery vehicle, confers a desired property to the
XX eukaryotic cell, or encodes a therapeutic proteinaceous substance or a
XX substance conferring resistance for an antibiotic substance to a cell.
XX The method is useful for improving gene targeting efficiency. The method
XX is useful in the replacement of an active gene by an inactive gene, for
XX e.g. for the inactivation of genes controlling undesired side branches of
XX metabolic pathways, to increase the quality of bulk products such as
XX starch, or to increase the production of specific secondary metabolites
XX or to inhibit formation of unwanted metabolites, and also to inactivate
XX genes controlling senescence in fruits and flowers or that determine
XX flower pigments. The method is also useful for replacing an inactive gene
XX by an active gene. For e.g. the replacement of a defective p53 by an
XX intact p53. Many tumours acquire a mutation in p53 during their
XX development which renders it inactive and often correlates with a poor
XX response to cancer therapy. By replacing the defect p53 by an intact p53,
XX e.g. through gene therapy, conventional anti cancer therapy have better
XX changes of succeeding. The method is also useful for therapeutic
XX proteinaceous substance integration. A tumoricidal gene can be delivered
XX to a pre-determined site present only in e.g. proliferating cells, or
XX present only in tumour cells, e.g. to the site where a tumour antigen is
XX expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved
XX in non-homologous recombination
XX
XX Sequence 1292 AA;
XX
Query Match 6.9%; Score 104.5; DB 5; Length 1292;
Best Local Similarity 15.0%; Pred. No. 17;
Matches 64; Conservative 93; Mismatches 133; Indels 137; Gaps 13;
OY 7 EOTVEVKSALJETADGALDFYNYLDQV-----IPKTFDETIKE 46
DB 246 ETSVQKDAEVHNKEMMLKDLRLQDQVSIKTAERSTLFKEQQRQYALPBN-EDTIEE 304

QY 47 LSRFKQESQASVLVDGDKVL---LMSQDKYFEATQTVVEMCGVVTOL---LSAYILL 100
DB 305 LKWKSKFEERLALIGTKIRKMEREMVDTETTISLHNAKTNYMLEISKLOTEAEAHMLL 364
QY 101 FDEYNEKASAK-----DILIRILDGYNKLENAQKSLGSSQS--- 140
DB 365 ---KNERDSTIONIFFHYNLGNVPSTPFSTEVVLMITNRIKSRIGELMDLLDKKSNET 421
QY 141 -----FNNASGKLLALDSQ-----LTFDSEKSSY----- 165
DB 422 ALSTAWDCYMDANDRWKIEIAQKRAKDEIKMISKRIEIERDSFPEFISTVDVKQTD 481
QY 166 -----FQSQVDRIKEAYA-----GAAAGIVAGPGLIISY 196
DB 482 EREKQVQVELBRKTKQNSRGFGFESKIEQOEIYSLEHKITLNRERDVMAAGDAEDRLIT 541
QY 197 SI-----AAGVIEKGLIP-----ELNDRKAVQNFTSLSVTVKQANKDIDAAKKLA 244
DB 542 RIDECKDIRGVKGRLPPEKDMKREIVQALRSIEREYDDLSLKSREAKEVNMLOMKIQ 601
QY 245 TEIAAIGBIKTETETTRFYVDYD-----DLMLSLKGAAKKMTNCNEYQORHG 293
DB 602 EVNSLFGKNDKTESRKEYIESKLOAKQESVTIDAYPKLESADKDDRKREYNWANG 661
QY 294 KKTILE 299
DB 662 MRQWFE 667

RESULT 22

ABB64018
ID ABB64018 standard; protein; 685 AA.

AC ABB64018;
XX

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 18846.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08121.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX Disclosure; SEQ ID NO 18846; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 685 AA;

Query Match 6.9%; Score 104; DB 4; Length 685;
Best Local Similarity 20.9%; Pred. No. 8.3;
Matches 68; Conservative 51; Mismatches 143; Indels 64; Gaps 11;

QY 4 IFASQTVVVKSAETADGALDFYKYLDOVIPWK--TFDETIKELSRFKQESQASVL 61
DB 129 ISAENRKOLLESKNILDLKLESINKYDQELKNKIQLGTPVKLSARNASEAQATRAV 188
QY 62 VGDTKVLLMSQDKYFEATQTVVYEW---GVVTOLL SAYILLFDEYNEKKAQAKDILIR 118
DB 189 EQALKQL-----ETFYGMCDGNAVNLAERAKQRLWDMRSTFADIPNSLLN 236
QY 119 ILD-----DGVNKLNEAQSLLGSSQSFNAGSKLLALDSQLTNDPSEKSSYFQSO-- 169
DB 237 VVMKIKEEQLOHIMKLNESRGECTNDKPPPLNNYEVKLLKTKADM---LGLAAYFAAQKE 293
QY 170 -----VDRIRKEAYA--GAAAGIVAGPGLIISYIAAGV----- 202
DB 294 LELKEERPCQDYSVVDKLOSKVYRFGNISLGDENADELISDYLQVNMNFRSQNEF 353
QY 203 ----IEGKLIPELNDRLKAVQNFTSLSVTVKQANKDIDAAKLKLATEIAAIGIKTETE 258
DB 354 LREQIE-QLRLELDAGAKOLEHDLKLG-SVKQYVGDINSSINRIQQDMVQLSQUIKEKIL 411
QY 259 TTRFYVD--YDDLMLSLKGAAKKMI 282
DB 412 FSRNMKNLLDDMQAATQKQNAKSQL 437

RESULT 23

AAU35203

ID AAU35203 standard; protein; 722 AA.

AC AAU35203;

DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation protein #490.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.

XX Enterococcus faecalis.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

XX 23-MAY-2000; 2000US-0206848P.

XX 26-MAY-2000; 2000US-0207727P.

XX 23-OCT-2000; 2000US-0242578P.

XX 27-NOV-2000; 2000US-0253625P.

XX 22-DEC-2000; 2000US-0257931P.

XX 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITEA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS53062.

XX New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids.

Example 3; SEQ ID NO 10796; 51lpp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 722 AA;

Query Match 6.9%; Score 104; DB 4; Length 722;

Best Local Similarity 19.8%; Pred. No. 8.9; Mismatches 109; Indels 108; Gaps 12; Matches 66; Conservative 51;

16 ALETADGALDFYKYLQDVIPWKTFTETIKELSRFKQY-----SQE 57
122 AVKGNNDKISYDDLKGTAAKVGTSANFLEKNKEKYDYTIKNFDDATGLYKALENGE 181
58 ASVLVGDIXVL-----LMSQD-----KYFEATQTVVWCG 88
182 ADAIVDDYFVLGVAVKNGOKLQVGDKETGSSYGFVAVKKGQNPelikFNAGLKNLKNK 241
89 VVTLSSAVILLFDEYNEKKAQAQKILIRLDDGVNKLNEAKSLLGSSQS-----FN 142
242 TYDKILNNVLTGDEINTQDAGEQ-----MKITPKKVKYVIASDSTFAPFEFQ 290
143 NASGKLALDSQLTNDFSEKSY-----FQSQVDRIRKAYAGAAAGIVAGPFGLII 194
291 NAQGDYVGIDVDLIVKRAELQGFTEFKFIFGSSAVQAVE-----SQADGMVAG----- 340
195 SYSIAGVIEGKLIPELNDRLKA-----VQNFTSLSVTVKQAN---KQIDAAK-----LK 242
341 -----WTITDRKKAQAFDSVPYFDSGIGQIAVKGNDKIKSYDDLKGGKVGK 387
243 LATEIA-AIGEIKTETETTRFFVYDDLMMLSLK 275
388 IGTESADFLKKNKKYDYSIKYLDTTDALYSALE 421

RESULT 24

ADA32997
ID ADA32997 standard; protein; 815 AA.

XX ADA32997;

XX 20-NOV-2003 (first entry)

XX Acinetobacter baumannii protein #158.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX plant biocontrol agent.

XX Acinetobacter baumannii.

XX US6562958-B1.

XX 13-MAY-2003.

XX

PF 04-JUN-1999; 99US-00328352.

XX 09-JUN-1998; 98US-0088701P.

PR (GENO-) GENOME THERAPEUTICS CORP.

XX Breton G, Bush D;

XX WPI; 2003-576092/54.

XX N-PSDB; ADA28871.

XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

XX A. baumannii and other Acinetobacter species in a sample, in screening

XX compounds for the ability to interfere with the A. baumannii life cycle

XX or to inhibit A. baumannii infection, and as biocontrol agents for

XX plants. The present sequence represents the amino acid sequence of an A.

XX baumannii protein.

XX Example; SEQ ID NO 4284; 328pp; English.

PS The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

XX A. baumannii and other Acinetobacter species in a sample, in screening

XX compounds for the ability to interfere with the A. baumannii life cycle

XX or to inhibit A. baumannii infection, and as biocontrol agents for

XX plants. The present sequence represents the amino acid sequence of an A.

XX baumannii protein.

XX Example; SEQ ID NO 4284; 328pp; English.

PS The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

XX A. baumannii and other Acinetobacter species in a sample, in screening

XX compounds for the ability to interfere with the A. baumannii life cycle

XX or to inhibit A. baumannii infection, and as biocontrol agents for

XX plants. The present sequence represents the amino acid sequence of an A.

XX baumannii protein.

XX Example; SEQ ID NO 4284; 328pp; English.

PS The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

XX A. baumannii and other Acinetobacter species in a sample, in screening

XX compounds for the ability to interfere with the A. baumannii life cycle

XX or to inhibit A. baumannii infection, and as biocontrol agents for

XX plants. The present sequence represents the amino acid sequence of an A.

XX baumannii protein.

XX Example; SEQ ID NO 4284; 328pp; English.

PS The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

XX A. baumannii and other Acinetobacter species in a sample, in screening

XX compounds for the ability to interfere with the A. baumannii life cycle

XX or to inhibit A. baumannii infection, and as biocontrol agents for

XX plants. The present sequence represents the amino acid sequence of an A.

XX baumannii protein.

XX Example; SEQ ID NO 4284; 328pp; English.

PS The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

XX A. baumannii and other Acinetobacter species in a sample, in screening

XX compounds for the ability to interfere with the A. baumannii life cycle

XX or to inhibit A. baumannii infection, and as biocontrol agents for

XX plants. The present sequence represents the amino acid sequence of an A.

XX baumannii protein.

XX Example; SEQ ID NO 4284; 328pp; English.

PS The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

XX for diagnosing a bacterial disease, as components of antibacterial

XX vaccines, as targets for antibacterial drugs, to detect the presence of

KW Streptococcus; GAS; CBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW anti-inflammatory; infection; vaccine; meningitis; gene therapy.
XX Streptococcus pyogenes.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

XX 24-NOV-2000; 2000GB-00028727.

XX 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;

XX Tettelin H;

XX WPI: 2002-352536/38.

XX N-PSDB; ABN69961.

XX New Streptococcus protein for the treatment or prevention of infection or

XX disease caused by Streptococcus bacteria, such as meningitis, and for

XX detecting a compound that binds to the protein.

XX Claim 1; Page 3916; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus (Streptococcus agalactiae) or group A streptococcus (GAS

XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

XX the specification. The proteins have antibacterial and anti-inflammatory

XX activity. (i), nucleic acids encoding (i), ABN6044-ABN71526 and

XX antibodies that bind (i) are used in the manufacture of medicaments for

XX the treatment or prevention of infection or disease caused by

XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX Nucleic acids encoding (i) are used to determine whether a compound binds to

XX biological sample. (i) is used to determine whether a compound binds to

XX (i). A composition comprising (i) or a nucleic acid encoding (i), may be

XX used as a vaccine or diagnostic composition. The disease caused by

XX Streptococcus that is prevented or treated may be meningitis. Nucleic

XX acid encoding (i) may be used to recombinantly produce (i) and may be

XX used in gene therapy. Antibodies to (i) are used for affinity

XX chromatography, immunoassays, and distinguishing/identifying

XX Streptococcus proteins

XX Sequence 541 AA;

XX Query Match 6.9%; Score 103.5; DB 5; Length 541;

XX Best Local Similarity 21.6%; Pred. No. 6.7;

XX Matches 71; Conservative 51; Mismatches 142; Indels 65; Gaps 14;

QY 26 FYNKLDQVPMKTFD--ETIKELSRFK-QEYSQASVILVG-----DIKVLMDSDQK 75

DB 49 FYQYADK-----EKIDLSEARSRASELDSAYOKKAKELVAKELRREGKIVTRDDPTH 104

QY 76 YFEATQTVYEWCGVVTOLLSAVILLFDEYNEKASAKQDKILIRILLDDGWNKLNKQXSL 135

DB 105 QENADMISYN-LAMKTNALRLRLNIDLEMOELANGHEKLTCKFLDEGYRKEFQAGLL 163

QY 136 SSSQSFNNASGKLLALDLSOLTNDP--SEKSSYFQSDVRIRKEAYAGAAIVAGPFGLI 193

DB 164 GLSVA-SQASVKSIA-DAVINANFKGAKWSNIDWQDKLSRIISQSVQSAILKGNGLT 221

QY 194 I-----SYSIAAGVTEGKLIPELNDRLKAVONFTSLSVTVKQAKDI--DAA 239

DB 222 IARDIRREBFDVSASYAKRLAITEHARVQMEVGRLSMAENGAFMDILPEPKACDVCKDIA 281

QY 240 K-----LKLATEAALGEIKETETTRFYVDYD-----DLMLSL---- 273

DB 282 KHGPYHLDKWRIGENSPFPFHYCRCAIVGVDSGVATDR-QLDYNNRNNNIDLMKATQSF 340

QY 274 -----LKGAAKKWINTCNEYQQRHGKKTLL 297

DB 341 IINKDVRVSARKKVGTRYDFWQAQNTTKKI 369

RESULT 28

AAU38921

ID AAU38921 standard; protein; 660 AA.

XX AC AAU38921;

XX 11-SEP-2003 (revised)

DT 16-JAN-2002 (first entry)

XX C. trachomatis CT875 protein.

XX Chlamydia; sexually transmitted disease; PID; antibacterial;

XX pelvic inflammatory disease; antigen; trachoma; gynecological;

XX acute respiratory tract infection; atherosclerosis; male infertility;

XX coronary heart disease.

XX Chlamydia trachomatis; serovar E.

XX WO200181379-A2.

XX 01-NOV-2001.

XX 23-APR-2001; 2001WO-US013081.

XX 21-APR-2000; 2000US-0198853P.

XX 20-JUL-2000; 2000US-0219752P.

XX (CORI-) CORIXA CORP.

XX Bhatia A, Probst P, Stromberg EJ;

XX WPI; 2001-616771/71.

XX N-PSDB; AAS56995.

XX New polynucleotide for treating Chlamydia infections encodes a

XX polynucleotides containing an immunogenic portion of a Chlamydia antigen.

XX Claim 3; Page 205-207; 208pp; English.

XX The invention relates to isolated polynucleotide encoding at least a

XX partial Chlamydia protein which is an antigenic fragment, or the

XX complements, fragments, homologues and variants, and antibodies raised

XX against the antigenic proteins (or fragments). The nucleic acids,

XX proteins and antibodies are used to diagnose and treat Chlamydia

XX infections (e.g. a sexually transmitted disease, pelvic inflammatory

XX disease (PID), acute respiratory tract infection, trachoma,

XX atherosclerosis and coronary heart disease) in a patient, and in the

XX treatment of male infertility. The compounds of the invention are also

XX useful for detecting the presence of Chlamydia in a patient, and

XX stimulating and/or expanding T cells specific for a Chlamydia protein.

XX The present sequence represents a Chlamydia antigen. (Updated on 11-SEP-

XX 2003 to standardise OS field)

XX Query Match 6.9%; Score 103.5; DB 4; Length 660;

XX Best Local Similarity 21.7%; Pred. No. 8.7;

XX Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTVEVKSATITAGALDFYNYKL-----DQV-----IPWKTFF 40

DB 155 EVANNIKKALEAQKDTIDKLNKLVTLQNNKSLTEVLKTTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQESQASVILVGDIKVLMDSDQ--KYFEATQTVYEWCGVVTOLLSAYI 98

DB 215 DQIKDLER--QNTISYEA-VLTNAGEVIKASSEAGIKLQALQSIDV-AGDQSQ---AAV 267

isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 58016; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1020 AA;

Query Match 6.9%; Score 103.5; DB 6; Length 1020;
Best Local Similarity 19.8%; Pred. No. 15;
Matches 65; Conservative 41; Mismatches 90; Indels 133; Gaps 12;
3 GIFAETQVEVVK-----SAIETADG-----ALDFYKNKLDQVVPKTFDETIKLSR 49
159 GTF-EEAGVLKYQKQKKAQKLFETEDNLSRVQDIIELEEQLTPLAAQSEAAKEFLR 217
50 FQOEYSQ-EASVLVGDIKVLLMDSQDKYFEATQTVWCVGVTQLLSAYILLDFEYNEKK 108
218 LKETLTQTDVSLWVAEIKTAKD-----WDNKQ 245
109 ASAQKDILIRILDDGVNKLNEA---QKSLGSSQGFNNASGKLLALDLSLTNDFSEKSSY 165
246 AQLAKPNL-----ELGKLSGSIQOEISLAKQKENAQADRLIEKNQVLLDSEKLQ 299
166 FOSQVDRIRKEAYAGAAGIVAGPFGLIITSYIAAGVTEGKLIPELNRLKAVQNFPTSL 225
300 TEGQKD-----VLQERTKHTQSSQBY 321
226 SVTVKQAKDIDAAKLKATEAATGEIKETETTFYVDYDMLSLILKGAAK----- 280
322 QTSLEAAQK-----KVK-----HFELQESLMKAAAEKETEIQ 354
281 -----MINTCNEYQ--QRHGKTKTLEVPD 302
355 KAEANLIKQOELEKYQKSTKELLAELRD 383

RESULT 31
ABU23940
ID ABU23940 standard; protein; 1163 AA.

XX
AC ABU23940;
XX

19-JUN-2003 (first entry)
Protein encoded by Prokaryotic essential gene #9467.
Antisense; prokaryotic essential gene; cell proliferation; drug design.
Clostridium acetobutylicum.
W0200277183-A2.
03-OCT-2002.
21-MAR-2002; 2002WO-US009107.
21-MAR-2001; 2001US-00815242.
06-SEP-2001; 2001US-00948993.
25-OCT-2001; 2001US-0342923P.
08-FEB-2002; 2002US-00072851.
06-MAR-2002; 2002US-0362699P.
(ELIT-) ELITRA PHARM INC.
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
WPI; 2003-029926/02.
N-PSDB; ACA27810.
New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.
Claim 25; SEQ ID NO 51864; 1766pp; English.
The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 1163 AA;
Query Match 6.9%; Score 103.5; DB 6; Length 1163;
Best Local Similarity 21.7%; Pred. No. 18;
Matches 81; Conservative 55; Mismatches 142; Indels 95; Gaps 14;
1 MTGIFAEQTVVWKSIAETADG---ALDFYKNKLDQVVPKTFDETIK-----ELS 48

QY

Db 112 ITGDEVEVLEGAWSNEKCOBIIIGSLDDDFTRTV--VLPQGFSEFLKJGKERNNMLE 169
 Qy 49 RF--KQEYSQAS-----VLGGDIKVLMDSDQKYFEATQTVYEWCGVVT 91
 Db 170 RLFNQYQDELSFKLARKIRKEREKENVVGELAGYENINEDVLKERRELLKNNDFN 229
 Qy 92 QLLSAYILLFDEYNEKKAQAQDILI-----RILDDGVNKLAE---AQKSLIGSSQS--- 140
 Db 230 EASKEYLKAEEYNEGKEVWGQIEIEEKNRVRKDLMEKDEIDLKERRALGESSKVK 289
 Qy 141 -----FNASGKLALDSQLTN-----DFSEKSYFOSQVDRTRKEAYAGAA 182
 Db 290 FYIDNYENTLQIDILKEQILSRENTMKAISLEKEDMEKLSIAKNKEK----- 339
 Qy 183 AGIVAGPGLIISYSIAAGVTEGKLIPELNDRLKAVQNFFTSLSV-----TVKQAKND 235
 Db 340 ---ALPFMVKHIIILDAIKEKOLLDNKLEKKQLQKIEKLSLEANSKEELIKQNIKD 395
 Qy 236 IDAAKLKLATEIAAIGEIKTETTRFYVDYDDLMLSLKGAAKQMIN-----TCNEYQ 289
 Db 396 IDSLTKIQNLKESKIDNLKVPBE---YKNKINEGIFELRNVDEKLEKHKNLGLDCKDFQ 451
 Qy 290 ----QRHGKTYLL 298
 Db 452 VDFEKAWSKKEML 464

RESULT 32
 ADE63514
 ID ADE63514 standard; protein; 1976 AA.
 AC ADE63514;
 XX
 DT 29-JAN-2004 (first entry)
 DE Rat Protein Q9JLT0, SEQ ID NO 9458.
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX Rattus norvegicus.
 XX WO2003016475-A2.
 XX PD 27-FEB-2003.
 XX PF 14-AUG-2002; 2002WO-US025765.
 XX PR 14-AUG-2001; 2001US-0312147P.
 XX PR 01-NOV-2001; 2001US-0346382P.
 XX PR 26-NOV-2001; 2001US-0333347P.
 XX (GEO) GEN HOSPITAL CORP.
 PA (FARH) BAYER AG.
 XX Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX Claim 1; Page; 1017pp; English.
 XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1976 AA;

Query Match 6.9%; Score 103.5; DB 7; Length 1976;
 Best Local Similarity 21.9%; Pred. No. 37;
 Matches 61; Conservative 43; Mismatches 96; Indels 79; Gaps 10;
 Qy 10 VEVVKSATETADGALDFVNYLDQV-----IPWKTFTDETIKEL-----SRFKQEYSQE 57
 Db 1222 LKNNKQGLETDNKEACEVQLQVKAESHEKRRKLDQAQVELHAKVSEBGRDLRVELAEK 1281
 Qy 58 ASVL---VGDIKVLLMDSQDKYFEATQTVYEWCGVTVQLLSAYILLFDEYNEKKAQAQK 114
 Db 1282 ANKLQNELDNVSTLLEAEKKGKMFAK---DAAGLESQLODTBLLQEBETRQKLNLSR- 1337
 Qy 115 ILIRILDDGVNKLNEAOKSLGSSQSFNNASGKLALDLSQLTNDFSEKSSYFQSQVDRIR 174
 Db 1338 --IFQLSEKNSLQEQEE---EEEARKNLEKQVLAQSQLDAT-----KKKVD-- 1382
 Qy 175 KEAYAGAAAGIVAGPFGGLIISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVKQANK 234
 Db 1383 -----DLGTIEG-----LEBAKKLLK 1399
 Qy 235 DIDAAKLKLATEIAAIGEIKTETTRFYVDYDDLMLSL 273
 Db 1400 DVEALSQRLEKVLAYD--KLEKTKNRLQBELDLDLTVDL 1436

RESULT 33
 ADE63518
 ID ADE63518 standard; protein; 1976 AA.
 XX
 AC ADE63518;
 XX
 DT 29-JAN-2004 (first entry)
 DE Rat Protein Q9JLT0, SEQ ID NO 9462.
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX Rattus norvegicus.
 XX WO2003016475-A2.
 XX PD 27-FEB-2003.
 XX PF 14-AUG-2002; 2002WO-US025765.
 XX PR 14-AUG-2001; 2001US-0312147P.
 XX PR 01-NOV-2001; 2001US-0346382P.
 XX PR 26-NOV-2001; 2001US-0333347P.
 XX (GEO) GEN HOSPITAL CORP.
 PA

PA (FARB) BAYER AG.
XX Woolf C, D'urso D, Belfort K, Costigan M;
PI WPI; 2003-268312/26.
XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
PT Claim 1; Page; 1017pp; English.
PS
XX The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1976 AA;
Query Match 6.9%; Score 103.5; DB 7; Length 1976;
Best Local Similarity 21.9%; Pred. No. 37;
Matches 61; Conservative 43; Mismatches 96; Indels 79; Gaps 10;
QY 10 VEVVKSATFADGALFYNYKLDV-----IPKTFDETIKEL-----SRFKQYISOE 57
DB 1222 LEKNQGLTDTNDELACEVKVLOQVKAESGHRKFKKLDVQVQLHAKVSGDRLRVELAK 1281
QY 58 ASVL---VGDIKVLMDSDQKYFEATQTVYEWGCVVTQLLSAYILLDFEYNEKKAQAQD 114
DB 1282 ANKLQELNDVSTLLEAEKKGKFKAK---DAAGLESQLDQTQELLQETROKLNLSR- 1337
QY 115 ILIRILDGYNKLNKAEQKLLGSSQFNNAAGKLLALDSQLTNDPSEKSYFSQVDRLR 174
DB 1338 --IROLEEKNSLOEQE---BEEARKNLKQVLAQSLADT-----KKYVDD-- 1382
QY 175 KEAYAGAAGVPGFGLIISYSIAAGVIRGKLIPELNDRLKAVQNFFTSLSVTVQANK 234
DB 1383 -----DLGTLEG-----LEBAKKKLLK 1399
QY 235 DIDAAKLKLATEIAAIGETITETTRFYVDYDLMLSL 273
DB 1400 DVEALSQRLEBKVLAYD--KLEKTKNRLQQLDLDLTVDL 1436
RESULT 34
ID AAU14603
XX AAU14603 standard; protein; 5373 AA.
AC AAU14603;
XX
DT 24-OCT-2001 (first entry)

XX Novel bone marrow polypeptide #2.
DE Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
XX haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX Homo sapiens.
OS
XX WO200157187-A2.
PN
XX 09-AUG-2001.
PD
XX 05-FEB-2001; 2001WO-US003782.
PF
XX 03-FEB-2000; 2000US-00496914.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 30-NOV-2000; 2000US-0250683P.
XX (HYSE-) HYSEQ INC.
PA Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX WPI; 2001-488875/53.
DR N-PSDB; AAS22908.
DR
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
PT gene therapy.
PT
XX Claim 10; Page 198-208; 392pp; English.
PS AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used in
CC the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations or
CC deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptides. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein. The
CC nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins and
CC in assays to identify modulators of their expression and activity. The
CC anti-bone marrow protein antibodies and antagonists may also be used to
CC down regulate expression and activity. The antibodies may also be used as
CC diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be
CC used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID)
XX
SQ Sequence 5373 AA;
Query Match 6.9%; Score 103.5; DB 4; Length 5373;
Best Local Similarity 23.0%; Pred. No. 1.4e+02;
Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;
QY 29 KYLDQVTPKTFDETIKELSRFKQYISOEASVLGDIKVLMDSDQKYFEATQTVYEWCG 88
DB 1369 KYISDAL--RRLEEEKVVEEEKQHEVKELGVLGWSLTARNTQK---ATSETKEST 1423
QY 89 VVTQLLSAVILLDFEYNEKK---ASAQKDILIRILDGYNKLNKAEQKLLGSSQFN--- 142
DB 1424 DIEKAILEQQVLSSELTTKQGVSAIKTSQIFLAKHG-HKLSEKKEKQI--SEQLNALN 1480

QY 143 -----NASGKLLALDSQLTNDSEKSSYFQSQVDRIRKE-----AYAGAAAGIVAGP 189
 Db 1481 KAYHDLCDGSANQLQLOSLAHQTEQKT--LQKQNTCHQOLEDLCSWVGQERALAGH 1538
 QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAALKLA-TEIA 248
 Db 1539 QGRITQODLSA---LQKNQSDLKLDQDDIQNRATSFATVVKDIEGFMEENQTKLSPRELT 1595
 QY 249 AIGE 252
 Db 1596 ALRE 1599
 RESULT 35
 ADJ68935
 ID ADJ68935 standard; protein; 5373 AA.
 AC ADJ68935;
 XX
 DT 06-MAY-2004 (first entry)
 DE Human heat mitochondrial protein as a therapeutic target SeqID741.
 KW mitochondrial; human; screening assay; diabetes mellitus;
 KW Huntington's disease; osteoarthritis;
 KW Leber's hereditary optic neuropathy; LHON;
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
 KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
 KW osteopathic; ophthalmological; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO2003087768-A2.
 XX
 PD 23-OCT-2003.
 XX
 PF 04-APR-2003; 2003WO-US010870.
 XX
 PR 12-APR-2002; 2002US-0372843P.
 PR 17-JUN-2002; 2002US-0389987P.
 PR 20-SEP-2002; 2002US-0412418P.
 XX
 XX (MITO-) MITOKOR.
 PA (BUCK-) BUCK INST AGE RES.
 XX
 PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DE;
 XX
 DR WPI; 2003-845369/78.
 XX
 PT Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX
 PS Claim 1; SEQ ID NO 741; 180pp; English.
 XX
 CC This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, nootropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cytostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.

XX SQ Sequence 5373 AA;
 Query Match
 Best Local Similarity 6.9%; Score 103.5; DB 7; Length 5373;
 Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;
 QY 29 KYLDQVPMKTFDETIKELSRFQOEYASVLDGIKVLMLDSQDKYFEATQTYYWCG 88
 Db 1369 KYISDAL--RRLEBEKVVEBEKQEHVEKVKELLGWSTLARNTQKG--ATSETKEST 1423
 QY 89 VVTQLLSAYILLFDEYNEKK--ASAOKDILIRILDGWNKLENAQKSLGSSQSFN--- 142
 Db 1424 DIEKAILEQQVLSBELTTKEQVSEAIKTSQFLAKHG-HKLSEKEKQI--SEQLNALN 1480
 QY 143 -----NASGKLLALDSQLTNDSEKSSYFQSQVDRIRKE-----AYAGAAAGIVAGP 189
 Db 1481 KAYHDLCDGSANQLQLOSLAHQTEQKT--LQKQNTCHQOLEDLCSWVGQERALAGH 1538
 QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDAALKLA-TEIA 248
 Db 1539 QGRITQODLSA---LQKNQSDLKLDQDDIQNRATSFATVVKDIEGFMEENQTKLSPRELT 1595
 QY 249 AIGE 252
 Db 1596 ALRE 1599
 RESULT 36
 ADH85709
 ID ADH85709 standard; protein; 284 AA.
 AC ADH85709;
 XX
 DT 22-APR-2004 (first entry)
 DE Enterococcus faecalis polypeptide #189.
 KW Enterococcus faecalis infection; transcription regulatory element;
 KW antibacterial.
 OS Enterococcus faecalis.
 XX
 PN US6617156-B1.
 XX
 PD 09-SEP-2003.
 XX
 PF 13-AUG-1998; 98US-00134000.
 XX
 PR 15-AUG-1997; 97US-0055778P.
 XX
 PA (DOUC/) DOUCETTE-STAMM L A.
 PA (BUSH/) BUSH D.
 XX
 PI Doucette-Stamm LA, Bush D;
 XX
 DR WPI; 2003-895394/82.
 DR N-PSDB; ADH82304.
 XX
 PT New nucleic acid comprising a sequence encoding an Enterococcus faecalis
 PT polypeptide, useful for preparing a composition for diagnosing or
 PT treating E. faecalis infection.
 PS Disclosure; SEQ ID NO 3594; 193pp; English.
 XX
 CC The invention relates to Enterococcus faecalis polynucleotides and
 CC polypeptides. The invention also relates to a recombinant expression
 CC vector comprising a polynucleotide operably linked to a transcription
 CC regulatory element, a cell comprising a recombinant vector, a method for
 CC producing an E. faecalis polypeptide, an isolated nucleic acid comprising
 CC a sequence not given in the specification, a recombinant vector
 CC comprising the nucleic acid and a cell comprising the recombinant vector.
 CC The polynucleotides can be used to detect the presence of E. faecalis in

CC a sample. The sequences are useful for preparing a composition for
CC diagnosing or treating *Enterococcus faecalis* infection. This sequence
CC represents an *E. faecalis* polypeptide of the invention.

XX SQ Sequence 284 AA;
Query Match 6.8%; Score 103; DB 7; Length 284;
Best Local Similarity 21.0%; Pred. No. 3.2; Indels 100; Gaps 13;
Matches 67; Conservative 46; Mismatches 106;
QY 16 ALETADGALDFNFKYLDQVWPKVFDETIKEL-----SRFKQEQSQEASVLVDGIKVL--- 68
DB 6 AVKVGEGANFLEKNKEK-----VDYTIKNFDDATGLYKALENGEADAI VDDVPVLGYA 59
QY 69 -----LMSQD-----KYFEATQTVYVCGVVTQLLSAYILLFD 102
DB 60 VKNQGKQLQVGDKEGTSSGYFAVKKQNPFLIKKFNAGLKNLKGNGTYDKILNNYLATGD 119
QY 103 EYNEKKAQAQKDLIRILDDGVNKLNEAKSLGSSQS-----FNNASGKLALDSOLT 156
DB 120 ETNTQDAGEQ-----WKKITPKKEKVIASDSTFAPFFQNAQGYVGDVLDV 168
QY 157 NDFSEKSSV-----FQSQVDRIKRAYAGAAAGIVAGPFLIISYSIAAGVIEGKLI 208
DB 169 KRAELQGFTEVEKFGIFSSAQAQVE-----SQADGWAG-----M 205
QY 209 PELNRLKA-----VQNFSTLSVTVKQAN-----KDIDAAK-----LKLATEIA-AIGEIKT 255
DB 206 TITDDRKAFAFSPVYFDFSGIQAIVKGNDRKTSYDLDLKGKVGKIGTESADFLERKKK 265
QY 256 ETETRFYVDYDMLSL 274
DB 266 KYDYSIKYLDITDLYSAL 284

RESULT 37
AAG82283
ID AAG82283 standard; protein; 885 AA.
XX AC AAG82283;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:1660.
XX KW *Staphylococcus epidermidis* SRI strain; infection; diagnosis; vaccination;
XX OS endocarditis.
XX PN *Staphylococcus epidermidis*.
XX PD WO200134809-A2.
XX PF 17-MAY-2001.
XX PR 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX DR N-PSDB; AAH53133.
XX KW Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 18; Page 462; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC and (II), given in AAG81454 to AAG81120, from *Staphylococcus epidermidis*.
CC and (II) can have antibacterial activity and therefore can be used in

CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX for SEQ ID NO:4455 to 4464

XX SQ Sequence 885 AA;
Query Match 6.8%; Score 103; DB 4; Length 885;
Best Local Similarity 18.2%; Pred. No. 14;
Matches 53; Conservative 52; Mismatches 111; Indels 76; Gaps 8;
QY 42 ETIKELSRFKQEQSQEASVLVDGIKVLMSQDKYFEATQTVYVCG----- 87
DB 2 EPLKEEAIAKQKLSKEMEQQSDVIVTSDIDHYTDNQRLDERLNLHLSKQQAQKEGQQ 61
QY 88 GVTQTLLSAYILLFDEYNEKKAQAQKDLIRILDDGVNKLNEAKSLGSSQSFNNASGK 147
DB 62 AQINQLLOQY-----KGRQON-----DYDIEKLN---YELVKATENYEQLSGK 102
QY 148 LLALDSQLTND-----FSEKSSYFQSQVDRIKRAYAGAAAGIVAGPFLIISYSIAAG 201
DB 103 LNVLEERKKQSQETNARYEEELNLESQIDSINKKAQNE----- 142
QY 202 VIEGKLIPELNRL-----KAVQNFSTLSVTVKQANKDIDAANKLATEIAAIGEIKTET 257
DB 143 -----KLIAELKNKQKQLNKEVQELSLIYISDQHDKEKLEIKNSYTYLMSQSDVNNDI 198
QY 258 ETETRFYVDYDMLSLKGAQAKMINTCNEYQQ-----RHGKKTLLLEV 300
DB 199 RLEHTINEAKKSRDLSDRLVEAFNQLKDIOQNIQTQKEYQSSKKSMKV 250

RESULT 38
ADB67135
ID ADB67135 standard; protein; 961 AA.
XX AC ADB67135;
XX DT 04-DEC-2003 (first entry)
XX DE General vesicular transport factor p115 SEQ ID NO:155.
XX KW staged assembly; nanostructure; peptide nucleic acid; PNA;
KW structural reinforcement; aerogel; paper; plastic; cement;
KW tensile strength; identification marker; anti-counterfeiting marker;
KW enzyme support; catalyst support; assembly scaffold; nanowire;
KW nanocircuit; molecular sieve; molecular filter; biosensor.
XX OS *Bos taurus*.
XX PN WO2003072829-A1.
XX PD 04-SEP-2003.
XX PF 21-FEB-2003; 2003WO-US005390.
XX PR 21-FEB-2002; 2002US-00080608.
XX PA (NANO-) NANOFRAMES INC.
XX CC

PI Hyman PL, Goldberg EB;
XX WPI; 2003-721788/68.
XX Staged assembly of nanostructures, useful e.g. in biosensors or as
PT catalyst supports, using assembly units derived from peptide nucleic
PT acids.
XX
XX Disclosure; Page 58; 118pp; English.
XX
CC The present invention describes a method (M1) for the staged assembly of
CC a nanostructure using peptide nucleic acids (PNAs). M1 comprises: (a)
CC contacting a nanostructure intermediate (NSI) having at least one unbound
CC joining element (JE) with an assembly unit (AU) that comprises several
CC different JE where: (i) none of these JE can interact with itself or
CC other JE; and (ii) only one JE in AU and a single unbound JE in NSI are
CC complementary, so that AU becomes non-covalently linked to NSI to produce
CC a new NSI for use in subsequent cycles; (b) removing unbound AU; and (c)
CC cyclic repetition of (a) and (b) to form a nanostructure. The new feature
CC is that the complementary JE in at least one cycle are PNAs. Also
CC described are nanostructures formed from many AU, comprising different
CC JE, where at least one AU includes PNA. M1 is useful for producing
CC nanostructures with a very wide range of potential applications, e.g.
CC structural reinforcements (for aerogels, paper, plastics or cement,
CC particularly as long fibres to improve tensile strength); identification
CC (anti-counterfeiting) markers; enzyme or catalyst supports; assembly
CC scaffolds; for construction of nanowires or nanocircuits; size markers
CC for electron microscopy; molecular sieves and filters; substrates for
CC optical and other surface coatings; scaffolds for solubilising enzymes or
CC for trapping, protecting and delivering specific molecules; in high-
CC density computer memories; as artificial zeolite for absorbing ions from
CC water and for construction of new materials, including use in biosensors.
CC PNAs are more homogeneous than inorganic nanoparticles generally used to
CC form nanostructures, so will produce structures with predictable geometry
CC and stoichiometry. The present sequence represents a protein containing
CC coiled coil dimerisation sequences that can be used for structural
CC elements of assembly units, given in the exemplification of the present
CC invention.
XX
XX Sequence 961 AA;

Query Match 6.8%; Score 103; DB 7; Length 961;
Best Local Similarity 22.0%; Pred. No. 16;
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;
QY 11 EVKSAIETADGALDFYKNYL-----DQVTPWKTFDE-----TIKELSRFK 51
DB 639 EVKKTLEQHSIVTHYKNMIREQDLQLEELKQISTLKQNEQLQTAVTQVQSIQOQH 698
QY 52 QEYSQASVLVDIKVLL-MDSQDK--YFEATQTVYEWCGVVTQLLSAYILLFDEYNEKK 108
DB 699 DQYNL-----LKVQLGKDSHQGPYTDGAQ----MNGVQPEISR--LREEIEELK 743
QY 109 AS-----AKDILLIRLDGVNK--LNEAQKLLGSSQSFNASKLLALSQLTND 158
DB 744 SNRELQSLQAEKDSLIENLKSQSLSPGTNEQSSATAGDSEQIAELKQELATLKSQ-NS 802
QY 159 FSEKSYFQSOVDTRFRKAYAGAAAGIVAGPGLIISYSIAAGVTEGKL-----IPELN 212
DB 803 QSVETIKLQTEKQELQKTEAFKAPVPGSEVTIATKTTD--VEGRSLALQETKELK 860
QY 213 DRLKAVQNFFTSLSVTVKQANKDIDAALKLATEIAAIGEIKTEPETTRFYVDYDMLLS 272
DB 861 NEIKALSERTAIKQLDSSNSTI-----AILQNEKNKLEVDITDSKEQDQLLV- 910
QY 273 LLGAAKKNMINTCNEYQQ 290
DB 911 LLADQDQKIFSLKKNLKE 928
RESULT 39
ADL99406
ID ADL99406 standard; protein; 961 AA.

XX ADL99406;
AC 20-MAY-2004 (first entry)
DT Nanostructure assembly protein #57.
DE peptide nucleic acid; PNA; nanostructure.
XX
XX Synthetic.
XX US2003215903-A1.
XX 20-NOV-2003.
XX 21-FEB-2003; 2003US-00370685.
XX 21-FEB-2002; 2002US-00080608.
XX (HYMA)/ HYMAN P L.
XX (GOLD)/ GOLDBERG E B.
XX Hyman PL, Goldberg EB;
XX WPI; 2004-021840/02.
XX
XX Staged assembly of a nanostructure containing peptide nucleic acid
XX assembly units comprising contacting a nanostructure intermediate with an
XX assembly unit comprising different joining elements, and removing unbound
XX assembly units.
XX Disclosure; Page 43-44; 73pp; English.

The invention relates to staged assembly of a nanostructure comprising:
CC (a) contacting a nanostructure intermediate comprising at least one
CC unbound joining element with an assembly unit comprising different
CC joining elements; (b) removing unbound assembly units; and (c) repeating
CC steps (a)-(b) for a sufficient number of cycles to form a nanostructure,
CC where the assembly unit in at least one cycle comprises a peptide nucleic
CC acid. A single joining element of the different joining elements and a
CC single unbound joining element of the nanostructure intermediate are
CC complementary joining elements, where the assembly unit is non-covalently
CC bound to the nanostructure intermediate to form a new nanostructure
CC intermediate for use in subsequent cycles. The method for staged assembly
CC of a nanostructure further comprises: (i) capping the nanostructure with
CC at least one capping unit; and (ii) post-assembly conversion of specific
CC non-covalent interactions of complementary joining elements to covalent
CC linkages, where the linkages are stabilised. The nanostructure
CC intermediate comprises a surface bound initiator assembly unit. A first
CC assembly unit used in at least one cycle comprises at least one
CC structural element covalently linked to a first joining element
CC comprising a peptide nucleic acid, or a first structural element bound to
CC a second structural element to form a stable complex. The structural
CC element is covalently linked to the first joining element and to a second
CC joining element, comprising a peptide nucleic acid. The assembly unit
CC further comprises a functional element, comprising photoactive molecule,
CC photonic nanoparticle, inorganic ion, carbon nanotube, lectin or
CC chemiluminescent molecule. The assembly unit comprises sub-assembly units
CC that bind to each other to form a complex. The present sequence
CC represents a protein containing a coiled coil dimerisation sequence that
CC can be used as a structural element in the method of the invention.
XX
XX Sequence 961 AA;

Query Match 6.8%; Score 103; DB 8; Length 961;
Best Local Similarity 22.0%; Pred. No. 16;
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;
QY 11 EVKSAIETADGALDFYKNYL-----DQVTPWKTFDE-----TIKELSRFK 51
DB 639 EVKKTLEQHSIVTHYKNMIREQDLQLEELKQISTLKQNEQLQTAVTQVQSIQOQH 698
QY 52 QEYSQASVLVDIKVLL-MDSQDK--YFEATQTVYEWCGVVTQLLSAYILLFDEYNEKK 108

Db 699 DOYNL-----LKVQKDSQHOGPTDGAQ-----MNGVQPEISR---LREIEELK 743
Qy 109 AS-----AQKOLIRILDDGVNK--LNEAOKSLGSSQSFNNASGKLALDSQLTND 158
Db 744 SNRELQSLAEKDSLIENLKSQSPGTNEQSSATAGDSQIAELKQELATLKSQ--NS 802
Qy 159 FSEKSSYFQSDVRIRKEAYAGAGIVAGFPGLIISYIAAGVIEGKL-----IPELN 212
Db 803 QSVETIKLQTEKQELLQKTEAFKAPVPGSESVIATKTTD--VEGRSLALLOETKELK 860
Qy 213 DELKAVQNFFTSLSVTVKQANKDIDAAKLKATEIAAIGETETTRFVVDVDDLMIS 272
Db 861 NEIKALSBEERTAIEKQLOSSNSTI-----AILQNEKKLEVDITDSKKEQDILLV- 910
Qy 273 LLKGAAKKWMINTCNEYQQ 290
Db 911 LLADQDKIFSLKNKLKE 928
RESULT 40
ADJ69829
ID ADJ69829 standard; protein; 1583 AA.
XX AC ADJ69829;
XX DT 06-MAY-2004 (first entry)
XX DE Human heat mitochondrial protein as a therapeutic target-SeqID1635.
XX KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;
KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.
XX OS Homo sapiens.
XX PN WO2003087768-A2.
XX PD 23-OCT-2003.
XX PF 04-APR-2003; 2003WO-US010870.
XX PR 12-APR-2002; 2002US-0372843P.
XX PR 17-JUN-2002; 2002US-0389987P.
XX PR 20-SEP-2002; 2002US-0412418P.
XX PA (MITO-) MITOKOR
PA (BUCK-) BUCK INST AGE RES.
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
XX WPI; 2003-845369/78.
XX Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX Claim 1; SEQ ID NO 1635; 180pp; English.
XX This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,

CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nontropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX Sequence 1583 AA;
SQ Query Match 6.8%; Score 103; DB 7; Length 1583;
Best Local Similarity 19.9%; Pred. No. 30;
Matches 73; Conservative 55; Mismatches 89; Indels 150; Gaps 16;
Qy 7 EQTEVVVKSIAETADGALDFYKYLQVVPWK--FDETIKELSRFQEQYSQASVVLVD 64
Db 531 EQTIQ-----YNSLEKQKYNELTGLEETLKE---KDQNDQKLEKLMVQ 571
Qy 65 IKVLLMDSQDYFEATQTVYVWCGVTVQLLSAVI--LLFDYNEKKASAKDILIRILD-- 121
Db 572 MKVL---SEDK-----EVLSAEVKSLYEENN--KLSEKKQLSRDLEVF 610
Qy 122 -----DGVNKL--NEAOKSLGSSQSFNNASGKLIALD 152
Db 611 LSQKEDVILKEHITOLEKKQLQMWVEQDNLKLLENEQVKLFVKTQLY---GFLKEMG 666
Qy 153 SOLTNDSEKSSYFQSDVRIRKEAYAGAGIVAGFPGLIISYIAAGVIEGKL----- 207
Db 667 SEVSEDSKED-----VNVLOAVGSLAKINEEK 697
Qy 208 -----IPELNDRLKAVQNFTSLSVTVKQANKDIDAAKLKATEIAAIG----- 251
Db 698 NLAFQDEKVLLEKEIKCLQESVWQCBELSLRDYEQEKVLLRKELEIQSEKALQ 757
Qy 252 ----EIKTETETTFYVDYDMLSLKGAAGKWMINTCNEYQQRHGK-----KT 296
Db 758 SDLEMKNAEKTR--LENQNLLIQV-----EEVSQTCSEIHNKEKCFIKHENLKP 810
Qy 297 LLEVDPDI 303
Db 811 LLEQKEL 817
RESULT 41
ABB59344
ID ABB59344 standard; protein; 2056 AA.
XX AC ABB59344;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 4824.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DB N-PSDB; ABL03447.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 4824; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16351), expressed DNA
CC sequences (AB16176-AB16351) and the encoded proteins (AB16176-
CC AB16351). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2056 AA;

Query Match 6.8%; Score 103; DB 4; Length 2056;
Best Local Similarity 23.8%; Pred. No. 42;
Matches 65; Conservative 44; Mismatches 102; Indels 62; Gaps 13;

Qy 8 QTVEVKSATGALDFYNYLDQVPIPKTFDETIKLS-RPKQYSQ--EASVLVGD 64
Db 1296 ENLRKAKTVLEKAGTLEAEN--ADLATELRVNSRRQDRRRKQAESQTAEIQVKLAE 1353
Qy 65 IKVLLMDSQKPYFATQTVYEMCGVVTQLLSAYILLFDEYNEKKSAAQKILIRILDGV 124
Db 1354 IERARSELQEK---CTKLQQAENITNQL-----ERAEKASA-----AVKSASNM 1397
Qy 125 NKLINEAQKSL-----LGSSQSFNNAAGKLLALDSQLTNDFSEKSSY-----FQSQVD 171
Db 1398 SOLTEAQQLLEETFRQKLGSLSSKLRQIESEKEALQEQLEEDDEAKRNYERKLAETVTTQM 1457
Qy 172 RIRKEAVAGAAAGVAGPFGIIISYTAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQ 231
Db 1458 EIKKABEDA-----DLAKELEGK--KRLNKQTEALER-----QVKE 1493
Qy 232 ---ANKDIDRAKLATEIAAIGIKETETTR 261
Db 1494 LIAQNDRLDKSKKIQSELED-ATIELEAQRK 1525

RESULT 42
ABU27359
ID ABU27359 standard; protein; 472 AA.
XX
AC ABU27359;
XX
DT 19-JUN-2003 (first entry)
XX
XX Protein encoded by prokaryotic essential gene #12886.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Chlamydia trachomatis.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX

PI Wang L, Zamudio C, Malone C, Häselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA31229.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 55283; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 472 AA;

Query Match 6.8%; Score 102.5; DB 6; Length 472;
Best Local Similarity 23.1%; Pred. No. 6.8;
Matches 74; Conservative 42; Mismatches 105; Indels 99; Gaps 17;

Qy 25 DFYNKYLDQVPIPKTFDETIK-----ELSRFKQYEQEASVVLGDIKVLMD 72
Db 48 DFY--FLEHRIIFRVLDQAFKSDRDPMDPHLTGELKE-----RDQLNVIGGASYLI--- 96
Qy 73 QDKYFATQTVYEMCGVVTQLLSAYILLFDEYNE-----KKSAQKDLIRILD-- 121
Db 97 -----FLSEFAGT-----SAYI---BEYADIIRSKSILRKMIOAKDIEKKAEEP 139
Qy 122 -DGVNKLNEAQKSLIGSSQSFNNA-----SGKLLALDSQLTNDF-----SEKSSYFQSQVD 171
Db 140 RDVTTALDQNLFRISQTNFAPVYVLSKLGKLSSTKDKSFLALQEQEAFQASH 199
Qy 172 RIRKEAVAGAAAGV-----AGPFGIIISYTAAGVIEGKLIPELNDRLKAVONF-- 221
Db 200 DARIPTLSGPTTHFLDLDRMLNGFSPSNLI---LAARPAMGKTALN-----IVENCF 252
Qy 222 -----FTSLSVTVKQ-----ANKDIDRAKLATEIA-----AIGIKETETTR 261
Db 253 ESRLPVGIFSLMTVDQLIHRICSRSEVEAKKISVG-DISGRDFQRVSVVREMEHTL 311
Qy 262 FYVDYDDLMLSLKGRKKM 281
Db 312 LIDDYPLGKITDURARRM 331

RESULT 43
AAG46982
ID AAG46982 standard; protein; 746 AA.
XX AC AAG46982;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59165.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0128234P.
PR 06-APR-1999; 99US-0127462P.
PR 18-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140931P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.

Db

162

LERTQOANEALKAMDAERQOLRSANNKLRDPTIELRGLSQPKENKIETLQOSSLDDKQI

221

Qy

173

---IRKEAYAGAAAGIVAGPFGLIISYISIAAGVIEGKLIPELNDKRLKAVQNFSTLSVTV

229

Db

222

LEDLKKQLQAVEERKQIA-VTELSAKHOKNLEGLAEQVVDALSERDKAAET-ISSLQVLL

279

Qy

230

QKANKDI-----DAAKLKLATEI--AAIGBIKTETE

258

Db

280

AEKESKIAEMEAAATGEAARLRAAAATLKGELAHLKSENE

319

RESULT 44

AAG46981

ID AAG46981 standard; protein; 788 AA.

XX

AC AAG46981;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 59164.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

05-MAR-1999; 99US-0123180P.

PR

09-MAR-1999; 99US-0123548P.

PR

23-MAR-1999; 99US-0125788P.

PR

25-MAR-1999; 99US-0126264P.

PR

29-MAR-1999; 99US-0126785P.

PR

01-APR-1999; 99US-0127462P.

PR

08-APR-1999; 99US-0128234P.

PR

08-APR-1999; 99US-0128714P.

PR

16-APR-1999; 99US-0129845P.

PR

19-APR-1999; 99US-0130077P.

PR

21-APR-1999; 99US-0130449P.

PR

23-APR-1999; 99US-0130510P.

PR

28-APR-1999; 99US-0130891P.

PR

30-APR-1999; 99US-0131449P.

PR

30-APR-1999; 99US-0132048P.

PR

04-MAY-1999; 99US-0132407P.

PR

05-MAY-1999; 99US-0132484P.

PR

06-MAY-1999; 99US-0132485P.

PR

06-MAY-1999; 99US-0132486P.

PR

07-MAY-1999; 99US-0132487P.

PR

11-MAY-1999; 99US-0132863P.

PR

14-MAY-1999; 99US-0134256P.

PR

14-MAY-1999; 99US-0134218P.

PR

14-MAY-1999; 99US-0134219P.

PR

14-MAY-1999; 99US-0134221P.

PR

14-MAY-1999; 99US-0134370P.

PR

18-MAY-1999; 99US-0134768P.

PR

19-MAY-1999; 99US-0134941P.

PR

20-MAY-1999; 99US-0135124P.

PR

21-MAY-1999; 99US-0135353P.

PR

24-MAY-1999; 99US-0135629P.

PR

25-MAY-1999; 99US-0136021P.

PR

27-MAY-1999; 99US-0136392P.

PR

28-MAY-1999; 99US-0136782P.

PR

01-JUN-1999; 99US-0137222P.

PR

03-JUN-1999; 99US-0137528P.

PR

04-JUN-1999; 99US-0137502P.

PR

07-JUN-1999; 99US-0137724P.

PR

08-JUN-1999; 99US-0138094P.

PR

20-AUG-1999;

99US-0149722P.

PR

20-AUG-1999;

99US-0149723P.

PR

20-AUG-1999;

99US-0149929P.

PR

23-AUG-1999;

99US-0149902P.

PR

23-AUG-1999;

99US-0149930P.

PR

25-AUG-1999;

99US-0150566P.

PR

26-AUG-1999;

99US-0150884P.

PR

27-AUG-1999;

99US-0151065P.

PR

27-AUG-1999;

99US-0151066P.

PR

27-AUG-1999;

99US-0151080P.

PR

30-AUG-1999;

99US-0151303P.

PR

31-AUG-1999;

99US-0151438P.

PR

01-SEP-1999;

99US-0151930P.

PR

07-SEP-1999;

99US-0152363P.

PR

10-SEP-1999;

99US-0153070P.

PR

13-SEP-1999;

99US-0153758P.

PR

15-SEP-1999;

99US-0154018P.

PR

16-SEP-1999;

99US-0154039P.

PR

20-SEP-1999;

99US-0154779P.

PR

22-SEP-1999;

99US-0155139P.

PR

23-SEP-1999;

99US-0155486P.

PR

24-SEP-1999;

99US-0155659P.

PR

28-SEP-1999;

99US-0156458P.

PR

29-SEP-1999;

99US-0156596P.

PR

04-OCT-1999;

99US-0157117P.

PR

05-OCT-1999;

99US-0157753P.

PR

06-OCT-1999;

99US-0157865P.

PR

07-OCT-1999;

99US-0158029P.

PR

08-OCT-1999;

99US-0158232P.

PR

12-OCT-1999;

99US-0158369P.

PR

13-OCT-1999;

99US-0159293P.

PR

13-OCT-1999;

99US-0159294P.

PR

13-OCT-1999;

99US-0159295P.

PR

14-OCT-1999;

99US-0159329P.

PR

14-OCT-1999;

99US-0159330P.

PR

14-OCT-1999;

99US-0159331P.

PR

14-OCT-1999;

99US-0159637P.

PR

14-OCT-1999;

99US-0159638P.

PR

18-OCT-1999;

99US-0159584P.

PR

21-OCT-1999;

99US-0160741P.

PR

21-OCT-1999;

99US-0160767P.

PR

21-OCT-1999;

99US-0160768P.

PR

21-OCT-1999;

99US-0160770P.

PR

21-OCT-1999;

99US-0160814P.

PR

21-OCT-1999;

99US-0160815P.

PR

22-OCT-1999;

99US-0160980P.

PR

22-OCT-1999;

99US-0160981P.

PR

22-OCT-1999;

99US-0160989P.

PR

25-OCT-1999;

99US-0161404P.

PR

25-OCT-1999;

99US-0161405P.

PR

25-OCT-1999;

99US-0161406P.

PR

26-OCT-1999;

99US-0161359P.

PR

26-OCT-1999;

99US-0161360P.

PR

26-OCT-1999;

99US-0161361P.

PR

28-OCT-1999;

99US-0161920P.

PR

28-OCT-1999;

99US-0161992P.

PR

28-OCT-1999;

99US-0161993P.

PR

29-OCT-1999;

99US-0162142P.

Query Match

6.8%;

Score 102.5;

DB 3;

Length 746;

Best Local Similarity

22.5%;

Pred. No. 12;

Matches 63;

Conservative 51;

Mismatches 121;

Indels 45;

Gaps 12;

Qy

2

TGFAEQTEVVVKSAIETADGALDPYKYLQVLPWKTFDETIKELSRFKQYEQSEASVL

61

Db

62

TRVAAEQALEHLREAYSEADAKQYSSKFSQV--EQKLDQEIQRD---EKYADL----

112

Qy

62

VGDIKVLVMSQDKYFEATQTVYECGVVVTQLLSAYILLFDEYNE--KKASAQKDILIRI

119

Db

113

--DAKFTLHKR-----AKRIQE-----IQEKODLDARFVNETAERASSQHSMMQE

161

Qy

120

LDDGVNKLNEAKSLGSSQFNNASGKLALDLSQLTWDFSEKSYF-----OSQVDR---

172

Db

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 22-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145318P.
PR 27-JUL-1999; 99US-0145319P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158369P.
PR 12-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.8%; Score 102.5; DB 3; Length 788;
Best Local Similarity 22.5%; Pred No. 13;
Matches 63; Conservative 51; Mismatches 121; Indels 45; Gaps 12;
2 TGIFAEQTVVVKSAIETADGALDFYKYLQVTPWKTFTDIKELSRFKQESQASVL 61

db 104 TRVAAEQALEHLREYSEADAKSQYSSKFSQV--EQKLDQBIKRD---EKYADL----- 154
Qy 62 VGDIKVLLMDSQDKYFEATQTVYVCGVVTQLLSAYILLFDEYNB--KKAQAQDKILIRI 119
Db 155 --DAFTRLHR-----AKRIQE-----IQEKDLDARFVENVTAERASSQHSQMOE 203
Qy 120 LDDGVNKLNEAQKSLGSSQFNNAAGKLLALDSQLTNDFFSEKSYF-----QSQVDR--- 172
Db 204 LERTQQAQNEALKAMDAERQQLRSANNKLRDTIELRGSLQPKENKIETLQQLSLDKQI 263
Qy 173 ---IRKEAYAGAAAGIVACPGFLIISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSYTV 229
Db 264 LEDLKKQLQAVEERKQIA-VTELSAKHQNKLEGLAQVVDALSERDKAAET-ISSLQVLL 321
Qy 230 QKANKDI-----DAAKKLATEI--AAIGEIKTETE 258
Db 322 AEKESKIAEMEAAATGEARLARLRAAAETLKGELAHKSENE 361

RESULT 45

AAU14697
ID AAU14697 standard; protein; 5447 AA.
AC AAU14697;
XX
XX
XX 24-OCT-2001 (first entry)
XX
XX Novel bone marrow polypeptide #96.
XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX
XX Homo sapiens.
XX
XX WO200157187-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US003782.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 20-JUN-2000; 2000US-00598075.
XX
XX 19-JUL-2000; 2000US-00620325.
XX
XX 30-NOV-2000; 2000US-0250683P.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
XX Ren F, Drmanac RT;
XX
XX WPI; 2001-488875/53.
XX
XX N-PSDB; AAS23002.
XX
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
XX gene therapy.
XX
XX Claim 10; Page 124-127; 392pp; English.

AAU14602-AAU14794 represent novel bone marrow polypeptides of the
invention. The proteins and corresponding coding sequences may be used in
the prevention, diagnosis and treatment of diseases associated with
inappropriate bone marrow polypeptide expression. For example, to treat
disorders associated with decreased expression by rectifying mutations or
deletions in a patient's genome that affect the activity of the
polypeptides by expressing inactive proteins or to supplement the
patient's own production of the polypeptides. Additionally, the nucleic
acids may be used to produce the polypeptides, by inserting the nucleic
acids into a host cell and culturing the cell to express the protein. The
nucleic acid and its complementary sequences may also be used as DNA
probes in diagnostic assays to detect and quantitate the presence of

CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins and
CC in assays to identify modulators of their expression and activity. The
CC anti-bone marrow protein antibodies and antagonists may also be used to
CC down regulate expression and activity. The antibodies may also be used as
CC diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be
CC used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID)
XX
SQ Sequence 5447 AA;

Query Match 6.8%; Score 102.5; DB 4; Length 5447;
Best Local Similarity 23.0%; Pred. No. 1.7e+02;
Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;
Qy 29 KYLDQVIPWKTFTDIKELSRFQEQYSEASVLGDIKVLMLDSQDKYFEATQTVYVCG 88
Db 1421 KYISDAL--RRLEBEKVVEBEKQEHVKEKVLGWSVTLARNTQK--ATSETKEST 1475
Qy 89 VVTQLLSAYILLFDEYNEK--ASAQKDIILIRILDDGVNKLNEAQKSLGSSQSFN--- 142
Db 1476 DIEKAILLEQQVLSBELTTKQEVSEAIKASQIFLAKHG-HKLSEKEKKQI--SEQLNALN 1532
Qy 143 -----NASGKLLALDSQLTNDFFSEKSYFSQSQVDRIKE-----AYAGAAAGIVAGP 189
Db 1533 KAYHDLCDGSANQLQQLQSALAHQTEKT--LQKQNTCHQQLDLCWSVGQERALAGH 1590
Qy 190 FGLIISVSIAGVTEGKLIPELNDRLKAVQNFFTSLSYTVVQKANKDIDAQKLA-TEIA 248
Db 1591 QGRTTQQDLSA---LQKNQSDKLDLQDDIQNRATSFATVWKDIEGFMEENQTKLSPRELT 1647
Qy 249 AIGE 252
Db 1648 ALRE 1651

RESULT 46

AAE10039
ID AAE10039 standard; protein; 590 AA.
XX
XX AAE10039;
XX
XX 11-SEP-2003 (revised)
XX 29-NOV-2001 (first entry)
XX
XX N. meningitidis strain 2996 961c-741 fusion protein.
XX Heterologous expression; Neisserial protein; 961c-741 fusion protein.
XX
XX Neisseria meningitidis; 2996.
XX
XX WO200164920-A2.
XX
XX 07-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-IB000420.
XX
XX 28-FEB-2000; 2000GB-00004695.
XX 13-NOV-2000; 2000GB-00027675.
XX
XX (CHIR-) CHIRON SPA.
XX Arico'NB, Comanducci M, Galsotti C, Masignani V, Giuliani MM;
XX Pizza M;
XX WPI; 2001-557776/62.
XX N-PSDB; AAD17054.
XX
XX Heterologous expression for the expression of two or more Neisserial
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:07:04 ; Search time 41 Seconds
(without alignments)

711.065 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIFABQTEVVKSAIETA.....TCNEYQQRHGKTLLEVPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1472	97.5	305	2 AE0673	haemolysin HlyE [i
2	1381	91.5	305	2 C64864	hemolysin E - Esch
3	1366	90.5	305	2 E30838	hemolysin E [impor
4	1366	90.5	305	2 E85696	probable pore form
5	117	7.7	2139	2 T18296	myosin heavy chain
6	116.5	7.7	478	2 T12818	hypothetical prote
7	116.5	7.7	652	2 B59102	hypothetical prote
8	116	7.7	1023	1 LEECA	hemolysin A - Esch
9	115	7.6	1005	2 A64465	hypothetical prote
10	115	7.6	1093	2 AC1753	tail protein [bact
11	114.5	7.6	821	2 AC7087	hypothetical prote
12	113.5	7.5	587	2 A30394	bp22 protein homol
13	112.5	7.5	520	2 F70350	recombination prot
14	112.5	7.5	584	2 S75986	hypothetical prote
15	112.5	7.5	1024	2 S10056	hemolysin A - Esch
16	112.5	7.5	1098	2 B70232	hypothetical prote
17	112.5	7.5	2823	2 F87908	protein T22A3.8 [i
18	112.5	7.5	2823	2 T23064	hypothetical prote
19	112.5	7.5	3102	2 T43291	laminin alpha chai
20	111	7.4	329	1 VNU21	VSG expression sit
21	111	7.4	1295	2 T24587	hypothetical prote
22	110	7.3	1285	2 B72420	hypothetical prote
23	109	7.2	573	2 B97331	membrane associate
24	109	7.2	1492	2 T14652	protein J - Yersin
25	109	7.2	1545	2 T14966	phage lambda-relat
26	108.5	7.2	1959	1 A33977	myosin heavy chain
27	108.5	7.2	1999	1 S21801	myosin heavy chain
28	108	7.2	622	2 T22716	hypothetical prote
29	107	7.1	318	2 A45522	variant surface gl

30	106.5	7.1	1938	1 MWKW1	myosin heavy chain
31	106.5	7.1	2022	2 T43214	ovt1 protein - nem
32	106	7.0	1133	2 T22976	hypothetical prote
33	106	7.0	3187	2 JC5837	364K Golgi complex
34	105	7.0	2405	2 T08164	dynein alpha heavy
35	104.5	6.9	1115	2 T41342	probable coiled-co
36	104.5	6.9	1292	2 D84727	probable RAD50 DNA
37	104.5	6.9	1964	2 A59282	nonmuscle myosin I
38	104	6.9	640	2 T03754	hypothetical prote
39	104	6.9	1496	2 T05634	hypothetical prote
40	104	6.9	2334	2 S32920	cell wall-associat
41	103.5	6.9	1163	2 G97236	ATPase involved in
42	103.5	6.9	4588	2 T28667	dynein beta heavy
43	103	6.8	2017	1 A36014	myosin II heavy ch
44	103	6.8	2057	2 S61477	myosin II heavy ch
45	103	6.8	2155	2 AD2742	conserved hypotet
46	103	6.8	2155	2 C97523	hypothetical prote
47	102.5	6.8	379	2 D97198	methyl-accepting c
48	102.5	6.8	472	2 H81665	replicative DNA he
49	102	6.8	726	2 T44825	hypothetical prote
50	102	6.8	739	2 H75001	methyl-accepting c
51	102	6.8	927	2 AG1739	transmembrane prot
52	102	6.8	1039	2 S18199	myosin heavy chain
53	102	6.8	1147	2 T40866	cell polarity prot
54	101.5	6.7	472	2 G71503	probable replicati
55	101.5	6.7	1127	2 T28317	ORF MSV156 hypotet
56	101.5	6.7	1473	2 A35186	salivary agglutini
57	101	6.7	1475	2 T33318	hypothetical prote
58	101	6.7	1601	2 AB1730	hypothetical prote
59	100.5	6.7	540	2 T44967	gas-vesicle protei
60	100.5	6.7	566	2 S54091	hypothetical prote
61	100.5	6.7	861	2 D82814	ATP-dependent Clp
62	100.5	6.7	1009	2 E89792	hypothetical prote
63	100	6.6	451	2 T41722	probable gamma-glu
64	100	6.6	490	1 S71776	calcium-dependent
65	100	6.6	779	2 E97778	endopeptidase La (
66	100	6.6	927	2 AH1389	transmembrane prot
67	100	6.6	1278	2 T27925	hypothetical prote
68	100	6.6	1938	1 S06005	myosin alpha heavy
69	100	6.6	1938	2 T49464	alpha cardiac myos
70	99.5	6.6	579	2 G96966	methyl-accepting c
71	99.5	6.6	998	2 T00227	hemolysin A toxin
72	99.5	6.6	1066	1 A48689	kinasin-related pr
73	99.5	6.6	1961	1 A61231	myosin heavy chain
74	99.5	6.6	2094	2 S33124	tpz protein - huma
75	99.5	6.6	2819	2 A90551	conserved hypotet
76	99	6.6	481	2 B85575	probable glutamate
77	99	6.6	481	2 B90724	probable glutamate
78	99	6.6	520	2 S53393	cell division cont
79	99	6.6	539	2 F72288	methyl-accepting c
80	99	6.6	978	2 B89971	conserved hypotet
81	99	6.6	1079	2 T18356	membrane protein p
82	99	6.6	1188	2 G83960	chromosome segrega
83	99	6.6	1322	2 B71440	hypothetical prote
84	99	6.6	2052	2 C97038	phage-related prot
85	99	6.6	2346	2 T13829	tpz homolog - frui
86	98.5	6.5	558	2 S62458	vacuolar protein s
87	98.5	6.5	826	2 T33796	hypothetical prote
88	98.5	6.5	1992	2 A47297	myosin heavy chain
89	98	6.5	401	2 E83720	methyl-accepting c
90	98	6.5	403	2 H72216	outer membrane pro
91	98	6.5	633	2 T41332	casp homolog - fis
92	98	6.5	916	2 T51288	silencing protein
93	98	6.5	978	2 A70387	conserved hypotet
94	98	6.5	1040	2 T40859	silencing protein
95	98	6.5	1474	2 T18281	hypothetical prote
96	98	6.5	1939	1 A46762	myosin alpha heavy
97	97.5	6.5	760	2 C87029	hypothetical prote
98	97.5	6.5	804	2 B89961	leucyl-rRNA synthe
99	97.5	6.5	878	2 B84977	alanine-tRNA ligas
100	97.5	6.5	998	2 T41078	hemolysin - Escher

ALIGNMENTS

```
RESULT 1
AE0673
hemolysin hlyE [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain C
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0673
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AE0502; MUID:21534947; PMID:11677608
A:Accession: AE0673
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01758.1; PID:gl6502606; GSPDB:GN00176
C:Genetics:
C:Gene: STY1498
C:Superfamily: Escherichia coli hemolysin E

Query Match 97.5%; Score 1472; DB 2; Length 305;
Best Local Similarity 97.4%; Pred. No. 5.3e-91;
Matches 295; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MTGIFAEQTVVVKSAIETADGALDFYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 60
Db 3 MTGIFAEQTVVVKSAIETADGALDLYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 62

Qy 61 LVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 120
Db 63 LVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 122

Qy 121 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180
Db 123 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 182

Qy 181 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQANKDIDAAK 240
Db 183 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNRLKTVQNFTSLSVTVKQANKDIDAAK 242

Qy 241 LKLAETIAAIGIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 300
Db 243 LKLAETIAAIGIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 302

Qy 301 PDI 303
Db 303 PDV 305

RESULT 2
C64864
hemolysin E - Escherichia coli (strain K-12)
N:Alternate names: hemolysin-inducing protein
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: C64864
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64864
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-305 <BLAT>
A:Cross-references: GB:AE000216; GB:U00096; NID:gl1787417; PIDN:AACT4266.1; PID:gl1787430;
A:Experimental source: strain K-12, substrain MG1655
```

```
C:Genetics:
A:Gene: hlyE, hpr
C:Function:
A:Description: hemolytic activity
A:Note: pore formation
C:Superfamily: Escherichia coli hemolysin E
C:Keywords: cytolysis; cytotoxin; hemolysis; transmembrane protein
F:181-197/Domain: transmembrane #status predicted <TM>
F:123/Active site: Asp #status predicted

Query Match 91.5%; Score 1381; DB 2; Length 305;
Best Local Similarity 89.8%; Pred. No. 6.2e-95;
Matches 272; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MTGIFAEQTVVVKSAIETADGALDFYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 62

Qy 61 LVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 120
Db 63 LVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 122

Qy 121 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180
Db 123 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 182

Qy 181 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQANKDIDAAK 240
Db 183 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNKLSQVQFFTTLSNTVKQANKDIDAAK 242

Qy 241 LKLAETIAAIGIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 300
Db 243 LKLAETIAAIGIKETETTRFYVDYDDMLSLGAAKGMINTCNEYQQRHGKKTILEV 302

Qy 301 PDI 303
Db 303 PEV 305

RESULT 3
E90838
hemolysin E [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90838
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genon
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <HAY>
A:Cross-references: UNIPROT:Q9REB3; GB:BA000007; PIDN:BA35100.1; PID:gl3361141; GSPDB:G
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
C:Gene: ECs1677
C:Superfamily: Escherichia coli hemolysin E

Query Match 90.5%; Score 1366; DB 2; Length 305;
Best Local Similarity 88.8%; Pred. No. 6.2e-84;
Matches 269; Conservative 20; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MTGIFAEQTVVVKSAIETADGALDFYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNNKYLDQVVPKTFDETIKELSRFKQYSQEAASV 62

Qy 61 LVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 120
Db 63 LVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIRIL 122

Qy 121 DDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAG 180
```

Db 123 DDGTTKLNBAQKSLVSSQFNAGSKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAG 182
QY 181 AAAGVAGPFGLLIISYSIAAGVIEGKLIPELMDRLKAVONFPTSLSVTVKQANKDIDAAK 240
Db 183 AAGVAGPFGLLIISYSIAAGVIEGKLIPELMDRLKAVONFPTSLSVTVKQANKDIDAAK 242
QY 241 LKLAETIAAIGIKETETETTRFYVDYDMLSLKGAAGKMTNCNEYQORHGKKTLLFV 300
Db 243 LKLTETIAAIGIKETETETTRFYVDYDMLSLKGAAGKMTNCNEYQORHGKKTLLFV 302
QY 301 PDI 303
Db 303 PEV 305

RESULT 4
E85696
probable pore forming hemolysin hlyE [imported] - Escherichia coli (strain O157:H7, subsp. O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85696
R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: E85696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <STO>
A:Cross-References: UNIPROT:Q9REB3; GB:AB005174; NID:G12514879; PIDN:AA65033.1; GSPDB:G12514879
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: hlyE
C:Superfamily: Escherichia coli hemolysin E

Query Match 90.5%; Score 1366; DB 2; Length 305;
Best Local Similarity 88.8%; Pred. No. 6.2e-84;
Matches 269; Conservative 20; Mismatches 14; Indels 0; Gaps 0;
QY 1 MTGIFAEQTVVVKSAIETADGALDFYKYLQVDPKKTETIKELSRFKQESYQASV 60
Db 3 MTEIVADKTVVVKSAIETADGALDFYKYLQVDPKKTETIKELSRFKQESYQASV 62
QY 61 LVGDIKVLMDSQDKYFEATQTVVWCGVTVTQLLSAYILLDFEYNEKKAQKDLIRIL 120
Db 63 LVGNIKVLMDSQDKYFEATQTVVWCGVTVTQLLSAYILLDFEYNEKKAQKDLIRIL 122
QY 121 DDGVNKLNEAOKSLGSSQSFNAGSKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAG 180
Db 123 DDGTTKLNBAQKSLVSSQFNAGSKLLALDSQLTNDPSEKSSYFQSQVDKIRKEAYAG 182
QY 181 AAAGVAGPFGLLIISYSIAAGVIEGKLIPELMDRLKAVONFPTSLSVTVKQANKDIDAAK 240
Db 183 AAAGVAGPFGLLIISYSIAAGVIEGKLIPELMDRLKAVONFPTSLSVTVKQANKDIDAAK 242
QY 241 LKLAETIAAIGIKETETETTRFYVDYDMLSLKGAAGKMTNCNEYQORHGKKTLLFV 300
Db 243 LKLTETIAAIGIKETETETTRFYVDYDMLSLKGAAGKMTNCNEYQORHGKKTLLFV 302
QY 301 PDI 303
Db 303 PEV 305

RESULT 5
T18296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18296
R:Guillen, N.

submitted to the EMBL Data Library, February 1997
A:Reference number: Z18865
A:Accession: T18296
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-References: UNIPROT:Q07569; EMBL:L03534; NID:G1850912; PID:G1850913; PIDN:AA84804
C:Genetics:
A:Gene: mhca
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <WMO>
Query Match 7.7%; Score 117; DB 2; Length 2139;
Best Local Similarity 19.0%; Pred. No. 15;
Matches 75; Conservative 47; Mismatches 114; Indels 158; Gaps 13;
QY 7 EQTVVVKSAIETADGALDFYKYLQVDPKKTETIKELSRFKQESYQASVVLGDIK 66
Db 965 EVDITELNSQINTLNATVD-----KKDKTIAEQESIDEKEDETIKLKGDIK 1011
QY 67 VL-----LMSQDKYFEATQ 81
Db 1012 LLEEKODLEQDRADVSAKDDIAKLNKITECEDAKDEIAKLQEQLEDEENKKNLDTN 1071
QY 82 TVYEW---CGVTVQLLSAYILLDFEYNEKKAQKDLIRILDDGVNKL-----E 129
Db 1072 BLQQTQLKLGTEKSLAAQVAA-----TKKASDERDTLSQNLN--EKLTTKNTKTKRAD 1124
QY 130 AOKSLGSSQSF-----NNASGKLLALDSQLTNDPSEKSSYFQSQVDKIRK 175
Db 1125 LEKISGLKQDYEDLEDDKNIKIEGLRLNAQRIKELDDITKG-ADVQYLOKQ-----K 1178
QY 176 BAYAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELMDRLKAVONFPTSLSVTVKQANKD 235
Db 1179 EYESQ-----IAKQKEAIGNDVKNKKTIKEKELE 1212
QY 236 I-----DAKUK--LATEIAAIGIKETETETTRFYVDYDML----- 271
Db 1213 IQSLQEKLDTEVEBKEDAEKKKKEIKEMKALQEKENVENSSKNSKTEKDKKKLEDLNKDT 1272
QY 272 -----SLKLGAKKMTNCNEYQORHGK 294
Db 1273 QKLLDDMTADNEKUKAKADLEAQLNEVDQDNHEK 1306

RESULT 6
T12818
hypothetical protein yonD - Bacillus subtilis phage SPBc2
C:Species: Bacillus subtilis phage SPBc2
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: T12818; F69913
R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A:Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 prophage
A:Reference number: Z17583
A:Accession: T12818
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-478 <LAZ>
A:Cross-References: UNIPROT:O64067; EMBL:AF020713; NID:G3025478; PID:G3025532; PIDN:AA611
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardiniois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

```

QY 171 DRIRKEAVGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNF7SLSVTVK 230
DB 414 SRLKNH-----IGVYNKOLEKIEN-----ELE 435

QY 231 QANKQIDAIAKLIKLAT-EIAAIGIKTETETTRFYVDYDDI-----MLSLKGAAKKMIT 284
DB 436 ECKNKKIDNTKKQALAEFDKSNKKQOELESELVQLNKKIDELGKRHKHQRLEASQKALDE 495

QY 285 CNEYQQRHGKK 295
DB 496 AXEINKKLAEK 506

RESULT 8
LERCA
hemolysin A - Escherichia coli
C:Species: Escherichia coli
C>Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 09-Jul-2004.
C:Accession: A24433; I41280
R:Pelmllee, T.; Pellett, S.; Welch, R.A.
J. Bacteriol. 163, 94-105, 1985
A:Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.
A:Reference number: A24433; MUID:85234404; PMID:3891743
A:Accession: A24433
A:Molecule type: DNA
A:Residues: 1-1023 <FEL>
A:Cross-references: UNIPROT:P09983; GB:M10133; GB:M12863; NID:G146377; PIDN:AAA
A:Experimental source: strain J96, O4 serotype
R:Stanley P.; Packman, L.C.; Koronakis, V.; Hughes, C.
Science 266, 1992-1996, 1994
A:Title: Fatty acylation of two internal lysine residues required for the toxic
A:Reference number: A55387; MUID:95099325; PMID:7801126
A:Contents: annotation; lysine palmitoylation
A:Note: lysine modification is performed by the hlyC gene product
R:Haertlein, M.; Schiessl, S.; Wagner, W.; Rdest, U.; Kreft, J.; Goebel, W.
J. Cell Biol. 22, 87-97, 1983
A:Title: transport of hemolysin by Escherichia coli.
A:Reference number: I41280
A:Accession: I41280
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1,'T',3,'V',5,'T',7-44 <RES>
A:Cross-references: GB:M29173; NID:G146337; PIDN:AAA23957.1; PID:G146338
C:Genetics:
A:Gene: hlyA
C:Function:
A:Description: attacks blood cell membranes and causes cell lysis
C:Superfamily: hemolysin A; hemolysin A homology
C:Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tand
F:246-791/Domain: hemolysin A homology <HLYA>
F:723-851/Region: 9-residue repeat (G-G-X-G-[DN]-D-X-[LVIVFI]-X)
F:563,689/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match 7.7%; Score 116; DB 1; Length 1023;
Best Local Similarity 20.2%; Pred. No. 6.7;
Matches 68; Conservative 59; Mismatches 112; Indels 98; Gaps 11

QY 2 TGFAEQTVEVVK-----SAETADGALDFYKYLQDVIPIPKTPTDETIKLSRFKQE 53
DB 155 TALSSMKIDELIKKQKGGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFQQ 207

QY 54 YSQEASVL-----VGDIKVLMDSDQKYFEATQTVYEWGVVTVOLLISAYILL-FDEY 104
DB 208 LNLGSLVSNYTKHLNGVN-KLQNLPLNDIAGLDTV---SGILSAISAFILSNADAD 263

QY 105 NEKXASQAKOILIRLDD---GVNKLNEAQKSLIGSSQSFNNAASKLIALDSQLTNDPSE 161
DB 264 TGTKAAAGVELTTKVLGVNGVIGISQYIIAORAAQGLSTS-----YSTAAGVIEGKLIPELNDRLKA 217

QY 162 KSSYFQSDVRIKREAVAGAAAGIVAGPFGLLIIS-----YSTAAGVIEGKLIPELNDRLKA 217
DB 303 -----AAAGLIASVTVTAISPLSFLSIADKFRANKIEYSQRFKK 344

```


Nature 392, 353-358, 1998
 A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A;Reference number: A70300; MUID:98196666; PMID:9537320
 A;Accession: F70350
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-520 <AOF>
 A;Cross-references: UNIPROT:O66834; GB:AE000695; NID:G2983180; PIDN:AA06789.1; PID:G2983
 A;Experimental source: strain VPS
 C;Genetics:
 A;Gene: recN
 C;Superfamily: recN protein

Query Match 7.5%; Score 112.5; DB 2; Length 520;
 Best Local Similarity 20.1%; Pred. No. 4.7;
 Matches 59; Conservative 70; Mismatches 94; Indels 71; Gaps 13;

QY 10 VEVKSAIETADGALDFYNYKLDQVPIKTFDE---TIKLSRFQVSO---EASVL- 61
 Db 135 LEKVNSLRKKEQELFEFLRKEELIQKDYLFVRVEEIGISEEYEEELKNKANLIN 194
 QY 62 -VGDIKVLMDSDQKYFEATQVYWGCVVTTOLLS---AYILLFDEYNKAKASQKDI-- 115
 Db 195 NLEKVKKAVGESLYKLEGENSVYELGEIRKNLAKVESYSGKFSLEIKIANLEEVVE 254
 QY 116 -----LIRILDGVNKLNEAKSKLLGSSGFNNASGKL-----ALDSQLTN---DF 159
 Db 255 LYNLSKEEMPEISEEVEINEKLPRIORLEEKYKSPFELKEVEEIKKEELNSLNSVDF 314
 QY 160 SEKSSYFOSQVDRIRKEAYAGAAAGVAGPGLIISYIAAGVIEGKLPPELN-DRLKAV 218
 Db 315 KEEB-LREEVEKLRBE-----YDKLAEVSRDRKKA 345
 QY 219 QNFPTSLSVTVKQAKDIDAALKLATEIAAIGETETTRTFYVDYDDMLLS 272
 Db 346 EDLEERIEEILKEIN-LEKALKV-----EIK-ESEPTKYGKDKIEFLFS 388

RESULT 14
 S75986
 hypothetical protein - Synechocystis sp. (strain PCC 6803)
 A;Species: Synechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 A;Accession: S75986
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S75986
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-594 <KAN>
 A;Cross-references: UNIPROT:Q55486; EMBL:D64006; GB:AB001339; NID:G1001291; PIDN:BA01083
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Superfamily: arginine-tRNA ligase

Query Match 7.5%; Score 112.5; DB 2; Length 584;
 Best Local Similarity 23.2%; Pred. No. 5.4;
 Matches 76; Conservative 45; Mismatches 131; Indels 75; Gaps 16;

QY 3 GIPAEQTVVVKSAJETADGALD---FY---NKYLDQVPIKTFDETIKLSRFQKE 53
 Db 171 GMLITYLKEVPEALVTAD-ALDIDGLVTFYKQAKQRFQ-----DEQFRTSR----- 218
 QY 54 YSOEASVLVGDIK-----VLMDSDQKYFEATQVYWGCVVTO-----LISAYIL 99
 Db 219 -QAVVALQAGDAKSIAKAWQLCEQSRREF---QLIYDCLDITIERGSGSYNPFPLPGVVE 274
 QY 100 LPDEYNKAKASQKDIILIR-----ILDDGVNKLNEAKSKLLGSSQ-SFNNASKLIAL 151

Nature 392, 353-358, 1998
 A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A;Reference number: A70300; MUID:98196666; PMID:9537320
 A;Accession: F70350
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-520 <AOF>
 A;Cross-references: UNIPROT:O66834; GB:AE000695; NID:G2983180; PIDN:AA06789.1; PID:G2983
 A;Experimental source: strain VPS
 C;Genetics:
 A;Gene: recN
 C;Superfamily: recN protein

Query Match 7.5%; Score 112.5; DB 2; Length 520;
 Best Local Similarity 20.1%; Pred. No. 4.7;
 Matches 59; Conservative 70; Mismatches 94; Indels 71; Gaps 13;

QY 10 VEVKSAIETADGALDFYNYKLDQVPIKTFDE---TIKLSRFQVSO---EASVL- 61
 Db 135 LEKVNSLRKKEQELFEFLRKEELIQKDYLFVRVEEIGISEEYEEELKNKANLIN 194
 QY 62 -VGDIKVLMDSDQKYFEATQVYWGCVVTTOLLS---AYILLFDEYNKAKASQKDI-- 115
 Db 195 NLEKVKKAVGESLYKLEGENSVYELGEIRKNLAKVESYSGKFSLEIKIANLEEVVE 254
 QY 116 -----LIRILDGVNKLNEAKSKLLGSSGFNNASGKL-----ALDSQLTN---DF 159
 Db 255 LYNLSKEEMPEISEEVEINEKLPRIORLEEKYKSPFELKEVEEIKKEELNSLNSVDF 314
 QY 160 SEKSSYFOSQVDRIRKEAYAGAAAGVAGPGLIISYIAAGVIEGKLPPELN-DRLKAV 218
 Db 315 KEEB-LREEVEKLRBE-----YDKLAEVSRDRKKA 345
 QY 219 QNFPTSLSVTVKQAKDIDAALKLATEIAAIGETETTRTFYVDYDDMLLS 272
 Db 346 EDLEERIEEILKEIN-LEKALKV-----EIK-ESEPTKYGKDKIEFLFS 388

RESULT 14
 S75986
 hypothetical protein - Synechocystis sp. (strain PCC 6803)
 A;Species: Synechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 A;Accession: S75986
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S75986
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-594 <KAN>
 A;Cross-references: UNIPROT:Q55486; EMBL:D64006; GB:AB001339; NID:G1001291; PIDN:BA01083
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Superfamily: arginine-tRNA ligase

Query Match 7.5%; Score 112.5; DB 2; Length 584;
 Best Local Similarity 23.2%; Pred. No. 5.4;
 Matches 76; Conservative 45; Mismatches 131; Indels 75; Gaps 16;

QY 3 GIPAEQTVVVKSAJETADGALD---FY---NKYLDQVPIKTFDETIKLSRFQKE 53
 Db 171 GMLITYLKEVPEALVTAD-ALDIDGLVTFYKQAKQRFQ-----DEQFRTSR----- 218
 QY 54 YSOEASVLVGDIK-----VLMDSDQKYFEATQVYWGCVVTO-----LISAYIL 99
 Db 219 -QAVVALQAGDAKSIAKAWQLCEQSRREF---QLIYDCLDITIERGSGSYNPFPLPGVVE 274
 QY 100 LPDEYNKAKASQKDIILIR-----ILDDGVNKLNEAKSKLLGSSQ-SFNNASKLIAL 151

478 LKSENEIDNLKLSKETLSKFNDS-----LIQINELISTKNLQKQKMD----- 527
 QY 177 AVAGAAAGVAGPGLIISYIAAGVIEGKLPPELN-DRLKAVQNFPTSLSVTVKQAKD 235
 Db 528 -----LNNLNDNLKVVQDQKLIKNEETLKLKAE 556
 QY 236 IDAAKL---KLATEIAA-----IGEIKET-----ETTRFYVDYDDL----- 269
 Db 557 IDLSNSENDELKQITSKDDDEFKQWQSVYVEDEAKIRNAEVTNGLNEDIEDLKESKHL 616
 QY 270 --MLSLKGAACKMINTCNEYQQRHGKTKLLEVPDI 303
 Db 617 EETITELKNVHLENEC-ELEKQKFEXTSLELES 651

RESULT 12
 A90394
 bps2 protein homolog (bps2) [imported] - Sulfolobus solfataricus
 C;Species: Sulfolobus solfataricus
 C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
 C;Accession: A90394
 R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A;Description: Sulfolobus solfataricus complete genome.
 A;Reference number: A99139
 A;Accession: A90394
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-587 <KUR>
 A;Cross-references: UNIPROT:Q97WH8; GB:AE006641; NID:G13815540; PIDN:AAK42408.1; GSPDB:C
 C;Genetics:
 A;Gene: bps2

Query Match 7.5%; Score 113.5; DB 2; Length 587;
 Best Local Similarity 19.5%; Pred. No. 4.7; Indels 87; Gaps 14;
 Matches 64; Conservative 68; Mismatches 109; Indels 87; Gaps 14;

QY 3 GIPAEQTVVVK-----SAJETADGALDFYNYKLDQVPIKTFD 41
 Db 10 GITQYSELEKGVTSYEAPNAVYKTSLSRSLISLTSIKKAEEDLNVPADSYVEALD 69
 QY 42 ETI--KELSRPKQYQSEASVLVDIKVLMDSDQKYFEATQVYWGCVVTTOLLSA--- 96
 Db 70 NKLYRRRIKIRNGLGEENLIMDDRALLL---TYFSPENR-----LVQILSGDGN 119
 QY 97 ---YILLFDEYNKAKASQKDIILIRILDGVNKLNEAKSKLLGSSGFNN---ASOKLLA 150
 Db 120 VEFISTTSKINEIK-AKKELOKLTAEINARDELQK-----KYNNIREIOAKIRA 170
 QY 151 LDSQLTDFSEKSSYFOSQVDRIRKEAYAGAAAGVAGPGLIISYIAAGVIEGKLP 210
 Db 171 ID-----EEDIKLEKE--RESSNIVAK-----TYTIT--LFRQNKIN 206
 QY 211 LNDRLKAVQNFPTSLSVTVKQAKDIDAALKLATEIAAIGETETTRTFYVDYDD-- 268
 Db 207 ILNKIKVKQDELANLEFALKIEELQNKESKVSFDIKTQLEKEEINEKLLKLTNDRS 266
 QY 269 ---LMLSLKGAACKM-----INTCN 286
 Db 267 ELEIELKVLERVLEEVEVNSDRHLDTCN 294

RESULT 13
 F70350
 recombination protein RecN - Aquifex aeolicus
 C;Species: Aquifex aeolicus
 C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
 C;Accession: F70350
 R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Db 275 LLQE-----KDLIVEDNGAOCFLVDGFTNKDGBRLPLIVQKSDGGYNYATTDLAAL 325
Qy 152 DSQLTNDFSEKSSYP--OSQVDRIRKEAYAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIP 209
Db 326 NYRLNTDGAELIIVTDAGQANHFAQFPQVAEKAGILTDPTQV--HVPFGLVKGEDGK 382
Qy 210 ELNRLKAVQNFSTLSVTVQANKID-----AAKLKLAIEA---AIGEIK---T 255
Db 383 KLKTRAGTIRLKKOLLTEAVTRAQODLETRLTAERSETBFTKEVAQRVGIGAVKYADL 442
Qy 256 ETETTRFVVDYDDMLSLKGAACKMI 282
Db 443 SONRTSDVFSFDKMLAQGNTPYML 469

RESULT 15
S10056
hemolysin A - Escherichia coli plasmid pHy152
C:Species: Escherichia coli
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 04-Mar-2000
C:Accession: S10056
R:Hess, J.; Wells, W.; Vogel, M.; Goebel, W.
FEMS Microbiol. Lett. 34, 1-11, 1986
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with the nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison
A:Reference number: S07209
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1024 <HES>
A:Cross-references: ENBL:M14107
C:Genetics:
A:Genome: plasmid pHy152
C:Superfamily: hemolysin A; hemolysin A homology
C:Keywords: lipoprotein
E:247-792/Domain: hemolysin A homology <HYA>
F:564,690/Binding site: palmitate (lys) (covalent) #status predicted

Query Match 7.5%; Score 112.5; DB 2; Length 1024;
Best Local Similarity 20.5%; Pred. No. 11;
Matches 69; Conservative 55; Mismatches 116; Indels 97; Gaps 11;

Qy 2 TGIFAEQIVVVK-----SAETADGALDFYKYLVDVIPWKTFTETIKELSRPQOE 53
Db 155 TALSSMKIDELIKKQKSGNVSSSELAKASIELINQLVDVTASLN-----NNVNSFSQ 208
Qy 54 YSQEASVL-----VGDIKVLMDSDQKYFEATQTVYVCGVVTQLLSAYILL-FDEY 104
Db 209 LNTIGSVLNTKHLNGVN-KLQNLPLNDNIGAGLDTV---SGILSAISAFILSNADAD 264
Qy 105 NEKASAKQDILIRLDD---GVNKLNEAQSLLGSSQSFNNAQKLLALDSQLTNDPSE 161
Db 265 TRTKAAGVELTKVLGNVGKISQYIIAQRAAQGLSTS-----YSTAAGVIEGKLIPELNDRKA 217
Qy 162 KSSVFQSDVDRIRKEAYAGAAAGIVAGPFGLLIIS-----YSTAAGVIEGKLIPELNDRKA 217
Db 304 -----AAAAGLTASAVTTLAISPLSFLSIADKPKRANKIEYSQRPK 345
Qy 218 VONPFTSLSVTVQANKIDIAAKLKLATEIAAIGIKTETETTRFYVDYDDMLSLKGA 277
Db 346 LQYDGSLLAAFHKETGAGIDASLTISTIVLASVSSGISAATTSILV-----GAPVSALVGA 401
Qy 278 -----AKKMINTCNEYQORHGK 294
Db 402 VTGIISGILEASKQMPPEHVASKQADVIAEWEKKGK 438

RESULT 16
B70232
hypothetical protein BBG10 - Lyme disease spirochete plasmid G/lp28-2
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: B70232
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: B70232
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1098 <KLE>
A:Cross-references: UNIPROT:O50733; GB:AE000786; NID:g2690008; PIDN:AAC66075.1; PID:g2690
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 7.5%; Score 112.5; DB 2; Length 1098;
Best Local Similarity 20.6%; Pred. No. 13;
Matches 79; Conservative 60; Mismatches 122; Indels 123; Gaps 17;

Qy 10 VEVKSAIETADGALDFYK-----LDQVPMKTFETIKELSR 49
Db 552 VGVEKSVNEFNNSYDFVNEYQNLLEKESRERIIKTLPTDQVSALQKLNDINEKNK 611
Qy 50 -FKQEYQSEASVLVDIK--VLLMDSQDKYFEATQ-----TVYEWCGV 90
Db 612 AFVEKYGKSFETLNEARNQVVALEKQVNEYKLTALDRSFVEAKALQKEITDLEW---E 668
Qy 91 TOLLSAVILLDFEYNEKKASAKOILIRI-----LDD---GVNKLNEAOK----- 132
Db 669 TMLLPA-----KERASAEKQWASKIQAMKYKFVDEHKSQPKLNETWNTIKQYAE 719
Qy 133 -----SLGSSQSQFNNA-----SGKLLALDS-----QLTNDFSEKSSVFSQV 170
Db 720 KAQDTTKSLYDSMDGLNVFNKAPFKMDIAGFLNKDTGESIGEEFHNLLNGKDVNWGEGL 779
Qy 171 DRIKEAYAG-----AAAGIVAGPFGLLIISYSIAA-----GWIEGKLIPELNDRKAV 218
Db 780 EKMTQWYESWKTGKTAAGAVFGPWGEAVELNGLTDFVWGLKG---QEKARIKAI 835
Qy 219 QNFFTSLSVTVQANKIDIAAKLKLATEIAAI-----GEIKTETETTRFYVDYDDMLSL 273
Db 836 E-----KKDEDLLELEKSEVELKLEDRFDEIKRKEKLSDELDEYKTEIEF 885
Qy 274 LKGAAKMINTCNEYQOR-HGKKT 296
Db 886 LKQAQSKGQISGEFQKRLHDVQT 909

RESULT 17
F87908
protein T22A3.8 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: F87908; E87908
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999, Science 283, 2103, 1999; and
A:Accession: F87908
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2823 <STO>
A:Cross-references: UNIPROT:O45614; GB:chr_I; PIDN:CAA15432.1; PID:g3924779; GSPDB:GN0001
A:Accession: E87908
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2823 <ST2>
A:Cross-references: GB:chr_I; PIDN:CAB03385.1; PID:g3924881; GSPDB:GN00019; CESP:T22A3.8
C:Genetics:
A:Gene: T22A3.8
A:Map position: 1

60 VLVG------DIKVLMSQDK-----YFEATQTV-
::|| ::| :|| : ||:
1916 LVVKGRINRYKEVSNEIEKLVEAEADIAYSRNSIEKARSEELMNNFDEKKNNTMTLAEI 1975
84 ---YEWCGVVTTQLLSAYILLFDYEKK-----KASAKDILIRLLDDGVNKLN 128
1976 PDIVEQCQNITLL---YSOLIDEYDEYVTAGRHAEKLEVOAQK-IVDRFVDTTETEN 2031
129 -----EAQS-----LLGSSSQFNNASGKLIALDSQLTNDFSEKSXYFSQGVDRI 173
2032 PLKAHAYENIVALKNKATEAVDAASEAENVSKMLSGESGDANEES--LRSOLEKL 2089
174 RKEAVAGAAGIVAGPGLIIISYAAGVIE--GKIPELNDRLKAVONFFTSLSVTVKQ 231
2090 KNESLSN-----VDNSNAVKIVEELKKEKKDLTDRLGHNLNELKTSI----- 2131
232 ANKOIDAACKLATETAIAIGEKITTETTRFYDVYDDLMLSLLKGAAKM-INTCNEYQQ 290
2132 -----VRSLGVIKNEASS----WDDKHDMHSILKNGAKTAHERSANKVKE 2173
291 RHGKKT 296
2174 SEGKIT 2179

RESULT 19

T43291
laminin alpha chain - Caenorhabditis elegans
C;Species:Caenorhabditis elegans
C;Date:11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession:TJ43291
R;Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang,
submitted to the EMBL Data Library, June 1998
A;Description: Expression, function and evolution of laminin alpha chains.
A;Reference number: Z23197
A;Accession: T43291
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A;Residues: 1-3102 <ZHU>
A;Cross-references: UNIPROT:O45614; EMBL:AF074902; PIDN:AAC26793.1
C;Genetics:

A;Map position: 1
A>Note: lamal/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like hcp

Query Match 7.5%; Score 112.5; DB 2; Length 3102;
Best Local Similarity 19.9%; Pred.No. 50;
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

6 AQEIVEVV-----KSAIETADGALDPYNKYLDQVIPWKTFETTTELKELSRFKQEYSQEZAS 59
1862 ABEELKWVTAQKLNETIFEDLNKRIDVLEQMND-----YTETDYDVSCKDTADAERMS 1915
60 VLVG-----DIKVLMSQDK-----YFEATQTV----- 83
1916 LVVKGRINRYKEVSNEIEKLVEAEADIAYSRNSIEKARSEELMNNFDEKKNMTLAEL 1975
84 ---YEWCGVVTTQLLSAYILLFDYEKK-----KASAKDILIRLLDDGVNKLN 128
1976 PDIVEQCQNITLL---YSOLIDEYDEYVTAGRHAEKLEVOAQK-IVDRFVDTTETEN 2031
129 -----EAQS-----LLGSSSQFNNASGKLIALDSQLTNDFSEKSXYFSQGVDRI 173
2032 PLKAHAYENIVALKNKATEAVDAASEAENVSKMLSGESGDANEES--LRSOLEKL 2089
174 RKEAVAGAAGIVAGPGLIIISYAAGVIE--GKIPELNDRLKAVONFFTSLSVTVKQ 231
2090 KNESLSN-----VDNSNAVKIVEELKKEKKDLTDRLGHNLNELKTSI----- 2131
232 ANKOIDAACKLATETAIAIGEKITTETTRFYDVYDDLMLSLLKGAAKM-INTCNEYQQ 290
2132 -----VRSLGVIKNEASS----WDDKHDMHSILKNGAKTAHERSANKVKE 2173

QY 291 RHGKKT 296
DB 2174 SEGKT 2179

RESULT 20
NMUT21
VSG expression site-associated protein 221a precursor - Trypanosoma brucei
N:Alternate names: ESAG protein
C:Species: Trypanosoma brucei
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 09-Jul-2004
C:Accession: A03395
R:Cully, D.F.; Ip, H.S.; Cross, G.A.M.
Cell 42, 173-182, 1985
A:Title: Coordinate transcription of variant surface glycoprotein genes and an expression site-associated protein
A:Reference number: A90868; MUID:85254917; PMID:2861910
A:Accession: A03395
A:Molecule type: mRNA
A:Residues: 1-329 <CUL>
A:Cross-references: UNIPROT:P04478; GB:M11452; NID:G162072; PIDN:AAA30191.1; PID:G162073
C:Comment: The function of the ESAG proteins is not known but may be related to activation of the VSG expression site associated protein
C:Superfamily: VSG expression-site associated protein
C:Keywords: glycoprotein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-329/Product: VSG expression site-associated protein 221a #status predicted <MAT>
F:73,294,308/Binding site: carbohydrate (Ase) (covalent) #status predicted

Query Match 7.4%; Score 111; DB 1; Length 329;
Best Local Similarity 22.9%; Pred. No. 3,2;
Matches 48; Conservative 33; Mismatches 63; Indels 56; Gaps 9;

QY 92 QLLSAYILLFD-----EYNEKKASAKDILIRILDGVNKL-NEAOKS 133
DB 4 EIVELVLLFVTCTVDWMLQAGDCTRVADHKHAPTEAVCVLRCLSDALNKLKLYSEGEKK 63

QY 134 LLGSSQSFNNASGKLLALDSQLTNDPSEKSY-----FQSQVDRIRKEAYAGAAIV 186
DB 64 LLVTEEVANAS---LILD-DMEGRAGESSTVLSVIRGVMBEQTDRLK-----108

QY 187 AGPFLGIISYSIAAGVI---EGKLPELNDRLKAVQ-----NFFTSLSVTVKOA 232
DB 109 -----LISYNGMGNLVAKAGLFALEDLSKEVRKIPGALIKTNKYYSVAEIVRTV 162

QY 233 NKDIDAALKL-----ATEIAAIGEIKTE 256
DB 163 WEDVGEILWKETEAKCGSKVGEVGEIQTE 192

RESULT 21
T24587
hypothetical protein T0684.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T24587
R:Lloyd, C.
submitted to the EMBL Data Library, April 1996
A:Reference number: T24587
A:Accession: T24587
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1295 <WIL>
A:Cross-references: UNIPROT:Q22257; EMBL:Z70756; PIDN:CAA94789.1; GSPDB:GN00023; CESP:T0684
A:Experimental source: clone T0684
C:Genetics:
A:Gene: CESP:T0684.1
A:Map position: 5
A:Introns: 17/3; 218/2; 385/3; 481/3; 946/3; 1034/3; 1222/2

Query Match 7.4%; Score 111; DB 2; Length 1295;
Best Local Similarity 21.0%; Pred. No. 20;
Matches 65; Conservative 58; Mismatches 93; Indels 94; Gaps 13;

QY 6 AEQTEVVVKSATETADGALDFYNYKYLDQVIPWKTFTET-----IKELSRFQKQYSQASVL 61

DB 710 AEETNEKLSRLASSE-----EQILDKNQOESLIDDLKEKLSAESTNOELQVS 759
QY 62 VGDIKVILMDSQDKYFEATQTVYEWGCVVTTOLLSAYILLFDEYNEKKASA-----OKD1 115
DB 760 LEMLKIEVSNAQKWESE-----VLKESFEALQLELSASQVSVSVVDAVQEKDG 811

QY 116 LIRILDDGVNKLNEAOKS---LLGSS-----OSFN---NASGK- 147
DB 812 LLRLVDTLKLIETKESQAOLQOSSVEEIKQLQLDLQNFQKQNAEVLSEKLNKLNSSHKR 871

QY 148 -LLALDSQLTNDPSEKSSYFQSQVDRIRKEAYAGAAIVAGPFLGIISYSIAAGVIEGK 206
DB 872 DMVALASQL-BELOHKLKLVGSGQSVNKEEL-----IGAKIMNKE 910

QY 207 LIPELN---DRLKAVONFFTSLSVT-----VKQANKDIDAAKLKLATEIAAIG 251
DB 911 MVDELNAKLGALGEMELKSLSEVSEAKVORREELIAQVSKHRDQES-QLQLTLDLKL 969

QY 252 EIKTETETTR 261
DB 970 SAQHSSTETSR 979

RESULT 22
B72420
hypothetical protein TM0088 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: B72420
R:Nelson, K.E.; Claytton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: B72420
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1285 <ARN>
A:Cross-references: UNIPROT:Q9WXU3; GB:AE001695; GB:AE000512; NID:G4980569; PIDN:AAD35181
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0088

Query Match 7.3%; Score 110; DB 2; Length 1285;
Best Local Similarity 19.5%; Pred. No. 23;
Matches 69; Conservative 69; Mismatches 113; Indels 102; Gaps 20;

QY 1 MTGIPAEQT-----VEVVKSAIETADGALDFYNYKYLDQ-----VIPWKTFTETIKELSRF 50
DB 487 MYGVSMKIGNFLEFGARSSVENVEQL---KKLSBHTQFLKALKT--ENVEDLGRF 541

QY 51 -KQYISQASVILVDGIKVLMDSDQKYFEATQTVYEWGCVVTTOLLSAYILLFDEYNEK-K 108
DB 542 MKKYGVEFEYF-SSLKVMAMLSGKEE--ENVKAAEELQIISS--EERIIRFVKKTENVP 596

QY 109 ASAKDILIRILDDGVNKLNEAOKSLLGSSOSFNNASG---KLLALDSQLTNDPSEKSSY 165
DB 597 IDKAKNVVLQYSYSIEELGN-ELVVIGEREVEKADLLQKIFSEVEISRDFVKLPQSW 655

QY 166 FQSQVDRIRKEAYAGAAIVAGPFLGIISYSIAAGVI--EGKLPELNDRLKAVONFFT 223
DB 656 IDEQEKLE-----VVKNSAG--ITVEILDGVVYFEG-----685

QY 224 SLSVTVKQANKDIDAAKLKLATEIAAIGEIKTETETTRF-----YV 264
DB 686 -----TKENVEKAKELFSDIVEKLGEVRKB-ETVEFLEVNSFFVDEFINLSGKLYP 736

QY 265 D-----YDDLMLSLKGAAGKMINTCNEYQ---ORHGK-----KTLLEVP 301
DB 737 DVTCEFLDQLGLLVKGSSEAVEDLSSMYRFFERHOKIVKENVFDRLMLEVP 789

A:Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my
A:Reference number: A33977; MUID:90046668; PMID:2813355
A:Accession: A33977
A:Molecule type: mRNA
A:Residues: 1-1959 <SHO>
A:Cross-references: UNIPROT:P14105; GB:M26510; NID:9212382; PIDN:AAA48974.1; PID:9212383
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A:Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in
s.
A:Reference number: S06116; MUID:90032648; PMID:2806244
A:Accession: S06116
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 716-1008 <KAT>
A:Cross-references: GB:X17589
A:Note: this translation is not annotated in GenBank entry GGMHCFMHA, release 114
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A:Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate
A:Reference number: A43422; MUID:92381096; PMID:1512291
A:Accession: A43422
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1900-1959 <HOD>
A:Experimental source: brush border
A:Note: sequence extracted from NCBI backbone (NCBIP:111947)
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F:84-764/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1936/Domain: coiled coil #status predicted <COI>
F:837-1277/Region: S2
F:1278-1959/Region: light meromyosin
F:1937-1959/Domain: carboxyl-terminal <CBT>
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted

Query Match 7.2%; Score 108.5; DB 1; Length 1959;

Best Local Similarity 21.0%; Pred. No. 50;

Matches 66; Conservative 43; Mismatches 107; Indels 99; Gaps 11;

QY 7 EQTVVVKSAIE---TADGALDFNKKYLDQVIPWKTDFETIKE-----LSRFKQEYS 55

DB 1166 EQVTVVVKLTLEDEAKTHEAQIQEMRQKHSQAI--EELAEQLEQTKRVKANLEKAKQALE 1223

QY 56 QEASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFBYNEKKASAKQDI 115

DB 1224 SERAELSEVKKVLLQKQDA-----EHRKKVDAQLQE 1256

QY 116 L-----IRLLDGVNKLNEAKSLGSSQSFNNASGKLLALDS---QLTNDPFEKSSY 165

DB 1257 LQVKFTEGERVKTLEAERVKNLQVLEL-----DNVTGLNNGSDSKSLKAKDFSALESQ 1309

QY 166 FQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGVTEGK---LIPELNDRLKAVQNF 222

DB 1310 LQDTQELLQDET-----RLKLSFTKLTQTEDEKNALKEQLEEEEAENRLE 1356

QY 223 TSLSVTVKQA-----NKDIDAARKLKIATIAAIGETKETE 258

DB 1357 KQISVLQQQAVKARKMDDGLGCLAEAEAKKQLKDLLESLTQRYEEKIAAYD--KLEKT 1414

QY 259 TTRFVVDVDDMLSL 273

DB 1415 KTRLQQLDLDIAVDL 1429

RESULT 27

S21801

myosin heavy chain, neuronal [similarity] - rat

N:Alternate names: myosin II
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: S21801; PN0013; S18134
R:Sun, W.; Chantler, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A:Reference number: S21801; MUID:92235856; PMID:1569576
A:Accession: S21801
A:Molecule type: mRNA
A:Residues: 1-1999 <SUN>
A:Cross-references: UNIPROT:Q63731; EMBL:X62659
R:Sun, W.; Chantler, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral c
A:Reference number: PN0013; MUID:91151356; PMID:1998509
A:Accession: PN0013
A:Molecule type: mRNA
A:Residues: 1914-1998, 'I' <SU2>
A:Experimental source: brain
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F:84-763/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:541-575/Region: actin binding #status predicted
F:653-675/Region: actin binding #status predicted
F:836-1999/Domain: coiled coil #status predicted <COI>
F:836-1276/Region: S2
F:1277-1999/Region: light meromyosin
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:693,703/Active site: Cys #status predicted
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 7.2%; Score 108.5; DB 1; Length 1999;

Best Local Similarity 19.5%; Pred. No. 51;

Matches 66; Conservative 53; Mismatches 124; Indels 95; Gaps 13;

QY 7 EQTVVVKSAIE---TADGALDFNKKYLDQVIP--WKTDFET---IKELSRFKQEYSQE 57

DB 1165 EQVNLKTLLEDEAKTHEAQIQEMRQKHSQAVEELASQLEQTKRVKANLEKAKOTLENE 1224

QY 58 ASVLVGDITKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFBYNEKKASAKQDILI 117

DB 1225 RGLANEVKKVLLQGRD-----SEHRKKVQAEQLQELQ 1257

QY 118 RILDDGVNKLNEAKSLGSSQSFNNASGKLLALD---SOLTNDPS----- 160

DB 1258 VKFNEGERRVTELADKVTKLQVLELDNVTGLLSQSDSKSKLTKDFSALESQLDQTELLQ 1317

QY 161 -----EKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYSIAAGV 202

DB 1318 EENRQKLSLTKLQVEDEKNS--FREQLLEEEEAENKLEKQIA-----TLHAQV 1366

QY 203 IEGKLIPELNDRLKAVQNFFTSLSVTVKQANKNDAAKLKLATEIAAIGETKETEETTRF 262

DB 1367 ADMK--KQEDSVGCL-----TAEVKEKQLQDLEGLSQRHEEKVAYD--KLEKTKRL 1418

QY 263 YVDVDDMLSLKGAQKQKQNTCN--EYQQRHGKTKTLE 299

DB 1419 QOELDDLLVDL---DHQRQSACNLEKKQKKQKFDQLLAE 1452

RESULT 28

T27216

hypothetical protein F55C5.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T27216

R:Harris, B.

submitted to the EMBL Data Library, August 1996

136 --DALFRANKYTAIAEVRTVMDVKAIVE 163

RESULT 30

MWKW1

myosin heavy chain D [similarity] - Caenorhabditis elegans

N;Alternate names: myosin heavy chain I

N;Contains: myosin ATPase (EC 3.6.4.1)

C;Species: Caenorhabditis elegans

C;Date: 28-Feb-1986 #sequence revision 19-May-2000 #text_change 09-Jul-2004

C;Accession: T21193; T23973; S02772; A02993

R;McMurray, A.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z19388

A;Accession: T21193

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1938 <WIL>

A;Cross-references: UNIPROT:P02567; EMBL:Z71261; PIDN:CAA95806.1; GSPDB:GN00019; CESP:R06C7

A;Experimental source: clone F21C3

R;Gardner, A.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z19825

A;Accession: T23973

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1938 <W12>

A;Cross-references: EMBL:Z71266; PIDN:CAA95848.1; GSPDB:GN00019; CESP:R06C7.10

A;Experimental source: clone R06C7

R;Dibb, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.

J. Mol. Biol. 205, 603-613, 1989

A;Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gene

A;Reference number: S02771; MUID:89178677; PMID:2926820

A;Accession: S02772

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-376,'V',378-390,'V',392-576,'L',578-680,'I',682-1938 <DIB>

A;Cross-references: EMBL:X08065; NID:G6785; PIDN:CAA30854.1; PID:G6786

R;Karn, J.; Bremer, S.; Barnett, L.

Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983

A;Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy chain

A;Reference number: A93958; MUID:83273600; PMID:6576334

A;Accession: A02993

A;Molecule type: DNA

A;Residues: 24-93,'E',95-97,'R',99-376,'V',378-388,'GDV',392-407,'N',409-473,'G',475-576

C;Genetics:

A;Gene: myo-1; CESP:R06C7.10

A;Map position: 1

A;Introns: 23/3; 114/3; 229/1; 264/1; 320/1; 857/3; 1745/3; 1814/1; 1892/3

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle coiled coil

F;177-184/Region: myosin motor domain homology <NMOT>

F;177-184/Region: nucleotide-binding motif A (P-loop)

F;560-682/Region: actin binding #status predicted

F;764-778/Region: actin binding #status predicted

F;764-778/Region: actin binding #status predicted

F;846-1938/Domain: coiled coil #status predicted <COI>

F;846-1938/Region: S2

F;1161-1938/Region: light meromyosin

F;128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F;183/Binding site: ATP (Lys) #status predicted

F;700,710/Active site: Cys #status predicted

Query Match 7.18; Score 106.5; DB 1; Length 1938;

Best Local Similarity 17.38; Pred No. 67;

Matches 52; Conservative

8;

112; Indels 73; Gaps

51

5 FAEQTVVVK-----SAIFADGALDFYKYLQDVIPWKVDFEIKELSRFK

1315 FSSQLVEAKAAADELHERQBFACKNLEHLDQELLEEQINGK--DDIQQLSRIN 1372

52 QEYSQ-----EASVLVG--DIKVLIMDSQDKFEATQVVEWCGVVTQLLSAYILLFDEY 104

1432

```

140 QY 140 -SFNNAGKLLALD--SOLTNDSEKSSYFQSOVDRIKAEYAG-----AAAGIV 186
768 Db TELNLLTEKIRKVEIFERIKKNOELDHETARTADDLKQETNRNHLHLAKELEEARADIV 827
187 QY 187 AGPFGLIISYIAAGVIEGKLIPELNDRL-KAVQNFFTSLSVTVKQANKDIDAAK----- 240
828 Db A-----LNDRLAKMDANFKIKLDETIKSPADHETIKSRESK 864
241 QY 241 -----LKLATEIAAIGIKETETTRFRVYDDMLSLKGAAGKMWNTCNEVQOR 291
865 Db 865 SEKIIVKHETKIYEINKYRAELE--KLESDKDDL-----EKRIIGLQDELNEK 910

RESULT 32
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T22976; T23157
R/Lighting, J.
submitted to the EMBL Data Library, June 1994
A/Reference number: Z19645
A/Accession: T22976
A/Status: preliminary; translated from GB/EMBL/DBBJ
A/Molecule type: DNA
A/Residues: 1-1133 <WIL>
A/Cross-references: UNIPROT:Q21022; EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2.6
A/Experimental source: clone F59A2
R/Burton, J.
submitted to the EMBL Data Library, October 1995
A/Reference number: Z19700
A/Accession: T23157
A/Status: preliminary; translated from GB/EMBL/DBBJ
A/Molecule type: DNA
A/Residues: 1-1133 <WIL>
A/Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A/Experimental source: clone K01A11
C/Genetics:
A/Gene: CESP:F59A2.6
A/Map position: 3
A/Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1080/3; 1088/3; 1100/3; 1108/3; 1116/3; 1124/3; 1132/3; 1140/3; 1148/3; 1156/3; 1164/3; 1172/3; 1180/3; 1188/3; 1196/3; 1204/3; 1212/3; 1220/3; 1228/3; 1236/3; 1244/3; 1252/3; 1260/3; 1268/3; 1276/3; 1284/3; 1292/3; 1300/3; 1308/3; 1316/3; 1324/3; 1332/3; 1340/3; 1348/3; 1356/3; 1364/3; 1372/3; 1380/3; 1388/3; 1396/3; 1404/3; 1412/3; 1420/3; 1428/3; 1436/3; 1444/3; 1452/3; 1460/3; 1468/3; 1476/3; 1484/3; 1492/3; 1500/3; 1508/3; 1516/3; 1524/3; 1532/3; 1540/3; 1548/3; 1556/3; 1564/3; 1572/3; 1580/3; 1588/3; 1596/3; 1604/3; 1612/3; 1620/3; 1628/3; 1636/3; 1644/3; 1652/3; 1660/3; 1668/3; 1676/3; 1684/3; 1692/3; 1700/3; 1708/3; 1716/3; 1724/3; 1732/3; 1740/3; 1748/3; 1756/3; 1764/3; 1772/3; 1780/3; 1788/3; 1796/3; 1804/3; 1812/3; 1820/3; 1828/3; 1836/3; 1844/3; 1852/3; 1860/3; 1868/3; 1876/3; 1884/3; 1892/3; 1900/3; 1908/3; 1916/3; 1924/3; 1932/3; 1940/3; 1948/3; 1956/3; 1964/3; 1972/3; 1980/3; 1988/3; 1996/3; 2004/3; 2012/3; 2020/3; 2028/3; 2036/3; 2044/3; 2052/3; 2060/3; 2068/3; 2076/3; 2084/3; 2092/3; 2100/3; 2108/3; 2116/3; 2124/3; 2132/3; 2140/3; 2148/3; 2156/3; 2164/3; 2172/3; 2180/3; 2188/3; 2196/3; 2204/3; 2212/3; 2220/3; 2228/3; 2236/3; 2244/3; 2252/3; 2260/3; 2268/3; 2276/3; 2284/3; 2292/3; 2300/3; 2308/3; 2316/3; 2324/3; 2332/3; 2340/3; 2348/3; 2356/3; 2364/3; 2372/3; 2380/3; 2388/3; 2396/3; 2404/3; 2412/3; 2420/3; 2428/3; 2436/3; 2444/3; 2452/3; 2460/3; 2468/3; 2476/3; 2484/3; 2492/3; 2500/3; 2508/3; 2516/3; 2524/3; 2532/3; 2540/3; 2548/3; 2556/3; 2564/3; 2572/3; 2580/3; 2588/3; 2596/3; 2604/3; 2612/3; 2620/3; 2628/3; 2636/3; 2644/3; 2652/3; 2660/3; 2668/3; 2676/3; 2684/3; 2692/3; 2700/3; 2708/3; 2716/3; 2724/3; 2732/3; 2740/3; 2748/3; 2756/3; 2764/3; 2772/3; 2780/3; 2788/3; 2796/3; 2804/3; 2812/3; 2820/3; 2828/3; 2836/3; 2844/3; 2852/3; 2860/3; 2868/3; 2876/3; 2884/3; 2892/3; 2900/3; 2908/3; 2916/3; 2924/3; 2932/3; 2940/3; 2948/3; 2956/3; 2964/3; 2972/3; 2980/3; 2988/3; 2996/3; 3004/3; 3012/3; 3020/3; 3028/3; 3036/3; 3044/3; 3052/3; 3060/3; 3068/3; 3076/3; 3084/3; 3092/3; 3100/3; 3108/3; 3116/3; 3124/3; 3132/3; 3140/3; 3148/3; 3156/3; 3164/3; 3172/3; 3180/3; 3188/3; 3196/3; 3204/3; 3212/3; 3220/3; 3228/3; 3236/3; 3244/3; 3252/3; 3260/3; 3268/3; 3276/3; 3284/3; 3292/3; 3300/3; 3308/3; 3316/3; 3324/3; 3332/3; 3340/3; 3348/3; 3356/3; 3364/3; 3372/3; 3380/3; 3388/3; 3396/3; 3404/3; 3412/3; 3420/3; 3428/3; 3436/3; 3444/3; 3452/3; 3460/3; 3468/3; 3476/3; 3484/3; 3492/3; 3500/3; 3508/3; 3516/3; 3524/3; 3532/3; 3540/3; 3548/3; 3556/3; 3564/3; 3572/3; 3580/3; 3588/3; 3596/3; 3604/3; 3612/3; 3620/3; 3628/3; 3636/3; 3644/3; 3652/3; 3660/3; 3668/3; 3676/3; 3684/3; 3692/3; 3700/3; 3708/3; 3716/3; 3724/3; 3732/3; 3740/3; 3748/3; 3756/3; 3764/3; 3772/3; 3780/3; 3788/3; 3796/3; 3804/3; 3812/3; 3820/3; 3828/3; 3836/3; 3844/3; 3852/3; 3860/3; 3868/3; 3876/3; 3884/3; 3892/3; 3900/3; 3908/3; 3916/3; 3924/3; 3932/3; 3940/3; 3948/3; 3956/3; 3964/3; 3972/3; 3980/3; 3988/3; 3996/3; 4004/3; 4012/3; 4020/3; 4028/3; 4036/3; 4044/3; 4052/3; 4060/3; 4068/3; 4076/3; 4084/3; 4092/3; 4100/3; 4108/3; 4116/3; 4124/3; 4132/3; 4140/3; 4148/3; 4156/3; 4164/3; 4172/3; 4180/3; 4188/3; 4196/3; 4204/3; 4212/3; 4220/3; 4228/3; 4236/3; 4244/3; 4252/3; 4260/3; 4268/3; 4276/3; 4284/3; 4292/3; 4300/3; 4308/3; 4316/3; 4324/3; 4332/3; 4340/3; 4348/3; 4356/3; 4364/3; 4372/3; 4380/3; 4388/3; 4396/3; 44
```

```

192    SDELEAK-01QINRCKNCRZAEVLEENKXQSEYFERNNSDVEYFNQJLSDSDI1GCEHNGM1 210
OY      264   VYDDDLMSLLKGAAKKWINTCNEYQQRHGGKKTLLS 299
        | : : : | : | : |||
Db      249   AEAEEIVKQLLEEQAQSIIENLKDAENERNLKTALE 284

RESULT 33
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: JC5837
```

R.Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <OK>
A:Cross-references: UNIPROT:Q63714; DDBJ:D25543; NID:G516825; PIDN:BAA05026.1; PID:G5168
C:Comment: This protein plays a role in the formation and maintenance of the characteris
C:Superfamily: giantin
F:49-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predict
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 7.0%; Score 106; DB 2; Length 3187;
Best Local Similarity 22.2%; Pred. No. 1.4e+02;
Matches 78; Conservative 53; Mismatches 134; Indels 86; Gaps 16;

QY 6 AEQTV--EVVKGAIETADGALDFYKYLQVLPWKTFDTETTKELSRFKQEQYSQASVLVG 63
Db 1748 AEATLLANGAKPGVSTPSSHDDINNYLQDLQK---GRIAELEMERKQDRELSQTLN 1804

QY 64 DIKVL-----MDSQKYFE-----ATQTVYEWGVTOLLSAYILLFDEYNEKKAS 110
Db 1805 EKNALLTQISAKDSELKLEEEVAKINMLNQIQEELSERVTKL-----KETAE 1852

QY 111 AQKDILIRILDGWNKLENAQSKLLGSSQSFNNASGKLALDLSQTN-----DFSEKSS 164
Db 1853 EEKDDLEERL---MNQALENGISGNYQDVTDQIKNEQLESEMNLKRCVSELEBEKQ 1909

QY 165 YF---SQVD-RIRKAYAGAGIAGVAGPGLIISVIAAGVIEGKLPELNDRLKAVQ- 219
Db 1910 QLVKEKTKVSEIRKE-YMEKIQGAQKPGSKTHAKELQ-----ELLKEKQEVKQLQK 1962

QY 220 -----NFFTSLVTK-----QANKDIDAKKLATEI-----AAIGEIK----- 254
Db 1963 DCTRYLGRISALEKTVKALEFVHTESQKDLATKGNLAQAVEHHKKAQELSSFKILLDD 2022

QY 255 TETETTFRYVD-----YDDLMLSLKGAANKMINTCNEYQQRHKK 295
Db 2023 TQSEARVLADNLKLKELQSNKESQIKQKDEDLRLRLEQAEKHKE 2073

RESULT 34
T08164
dynein alpha heavy chain - Chlamydomonas reinhardtii (fragment)
C:Species: Chlamydomonas reinhardtii
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 02-Feb-2001
C:Accession: T08164
R:Mitchell, D.R.; Brown, K.S.
J. Cell Sci. 107, 635-644, 1994
A:Title: Sequence analysis of the Chlamydomonas alpha and beta dynein heavy chain genes.
A:Reference number: Z16302; MUID:94274778; PMID:8006077
A:Accession: T08164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2405 <MIT>
A:Cross-references: EMBL:I26049; NID:9415679; PIDN:AAA57316.1; PID:G603079
A:Experimental source: strain 21gr
C:Genetics:
A:Gene: ODA11
A:Note: Intron positions not resolved (incomplete sequence)
C:Superfamily: dynein heavy chain, ciliary
F:575-582/Region: nucleotide-binding motif A (P-loop)

Query Match 7.0%; Score 105; DB 2; Length 2405;
Best Local Similarity 21.0%; Pred. No. 1.1e+02;
Matches 75; Conservative 44; Mismatches 110; Indels 128; Gaps 14;

QY 38 KTFDETI---KELSRFKQYSQ-----EASVLVGDIKVLLMDSQDKY 76
Db 1793 KTFLELIKLYKNVLAARAKRANQNTERLENGHLKHLKQVADVDIILVEAKVKAVEHVKV 1852

QY 77 FE-----ATQTVYEWGVTOLLSAYILLFDEYNEKKASAKQDIL 116
Db 1853 ASANIPFAEQGVKEKVNENAAQVBEAKCVIAK-----EYSEKQASCEKDLA 1902

QY 117 IR--ILDDGWNKLENAQSKLLGSSQSF-----NNASGKLALDLSQL 155
Db 1903 AAEPLVAENAALETVTKDLGEAKSLKKPPGVDITAVVILLNNPKOKSWOAAQKL 1962

QY 156 TND---FSKSSYFQSDVR-----IRKEAY---AGAAAGIVAGPF 190
Db 1963 MNVDFLERVSKFSVIDAGQVARKTVDACRPLYALEWFNEEALCKKSAAGLCEWAV 2022

QY 191 GLIISYSIAAGV-----IEGKLIPELNDRLKAVQNFSTLSVTV-----KOANKD 235
Db 2023 NIITYDYVQVEPKQKQELAAANAKLEAVNTLAAVEEKVALLNAKVQEQYKEANDD 2082

QY 236 IDAA-----KLKLAIEAIAIGIKTETETTFRYVD-----YDDLMLSLKGA 278
Db 2083 KEAARESERCQKLELANRL--INALASEGERWALTVEQLRKSVEVLTGDMLLAA 2137

RESULT 35
T41342
probable coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41342
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21970
A:Accession: T41342
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1115 <LYN>
A:Cross-references: UNIPROT:O94488; EMBL:AL035076; PIDN:CAA22653.1; GSPDB:GN00066; SPDB:S
A:Experimental source: strain 972h-; cosmid c417
C:Genetics:
A:Gene: SPDB:SPCC417.07c
A:Map position: 1

Query Match 6.9%; Score 104.5; DB 2; Length 1115;
Best Local Similarity 20.6%; Pred. No. 44;
Matches 66; Conservative 60; Mismatches 127; Indels 67; Gaps 14;

QY 7 EQTVVVKSAJETAGALDFNFKYLDQVLPWKTFDEIKELSRFKQEQYSQASVLVGDIK 66
Db 500 EQSIETRLRYKQLQDIEEDYENKLMRMQQWR---EDVDQLQEVVEITQE---LQDTK 552

QY 67 VLLMDS---QDKYEATQTVYEWGVTOLLSAYILLFDE---YNEKKASAKQDILIRIL 120
Db 553 EVLKSSKESDDYBEV-----GKLTEAREIEKEFEKTIRENEESISLFEKEVEKLT 605

QY 121 D-----DGVNKLNEAQSKLLGSSQSFNNASGKLALDLSQTNDFSEKSSYFQSOV 170
Db 606 DEITQLSERVNDKCHEPDELQKRL---QTLSEENNAKEDSTKSTNLEQLKMTAEV 661

QY 171 DRIRKEAVAGAAGIVAGPFGLLISYSIAAGVIEGKL-----IPELNDRLKAVQNFST- 224
Db 662 DSLRKEENKQV-----IALKSELVKSNDNKLNLNQELISLNDQLKTEMESV 713

QY 225 -----LSVTVKQANKDIDAAKKLAT---EIAAIGIKTE-TETTFRYV---DY 266
Db 714 TTSKESLADYLSNLKXERHNDELDSLKKLRPEGLSSNLSLKSEIERNNQYVTLRENF 773

QY 267 DDLMLSLKGAANKMINTCN 286
Db 774 DSLQNAIME-TFDKQVTHCS 792

RESULT 36
D84727
probable RAD50 DNA repair protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: D84727
M:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84727
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1292 <STO>
A:Cross-references: UNIPROT:Q9SL02; GB:AE002093; NID:g4263721; PIDN:AAD15407.1; GSPDB:GN
C:Genetics:
A:Gene: At2g31970
A:Map position: 2
C:Superfamily: RAD50 protein

Query Match 6.9%; Score 104.5; DB 2; Length 1292;
Best Local Similarity 15.0%; Pred. No. 53;
Matches 64; Conservative 93; Mismatches 132; Indels 137; Gaps 13;

Qy 7 EQTEVVKSAETADGALDFNFKYLDQV-----IPWKTDETIKE 46
Db ETSQVKVDAEVRHNEKEMMLKDLRLQDQVSIKTAERSTLFKEQQRQYAAALPEEN-EDTIEE 304
Qy 47 LSRFKOYSQBASVLVGDIVKVL---LMSQDKYFEATQTVYEWCGVVTOL---LSAYILL 100
Db LKWKSKFEERLALLGCTKIRKMEREMVDTETISLHNAKTNMYLEISKLTQTEAEMHLL 364
Qy 101 FDEYNEKASAKQ-----DILIRLDDGVNKLNEAOKSLGSSQS--- 140
Db ---KNERDSTIONTFHYNLGNVPSTPSTEWNLNTRIKSRIGCELEMDLLDKKSNET 421
Qy 141 -----FNNSGKLLALDSQ-----LTNDSEKSSY----- 165
Db 422 ALSTAWDCYMDANDRWKSIQAQKRAKBIKMGISKRIBERDSFEFEISTVDVKQTD 481
Qy 166 -----FOSQVDRIKEAVA-----GAAAGIVAGPFLIISY 196
Db 482 EREKQVQVELRKTKQNSRGFESKIEQKHQHEIVSLEHKIKTLNRRDVRVMAQDAEDRLLT 541
Qy 197 SI-----AAGVIEGKLP-----ELNDRKAVQNPFTSLSVTVKQAKNDIDAAKKLA 244
Db 542 RIDECKDIRGVKGLPPEKDMKREIVQALRSIEREYDVLSLKSREAKEKVNMLQWKIQ 601
Qy 245 TEIAAIGIKETETTRFYVDYD-----DLMLSLKGAQAKMINTCNEYQQRHG 293
Db 602 EVNNSLFKHNKDTESRKRYIESKLQALKQESVTIDAYPKLJESAKDKDRDRKREYNMANG 661
Qy 294 KKTLL 299
Db 662 MRQMF 667

RESULT 37
A59282
nonmuscle myosin II heavy chain A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59282
R:Bhati,Dev, N.; Taira, M.; Nooruddin, H.; Adelstein, R.S.
Mech. Dev. 78, 33-36, 1998
A:Title: Differential expression of non-muscle myosin heavy chain genes during Xenopus e
A:Reference number: A59282; MUID:99077683; PMID:9858676
A:Accession: A59282
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1964 <BHA>
A:Cross-references: UNIPROT:Q93522; GB:AF055895; NID:g3660671; PIDN:AAC83556.1; PID:g366
A:Experimental source: cell line XTC
C:Superfamily: myosin heavy chain; myosin motor domain homology

F:84-764/Domain: myosin motor domain homology <MMO>
Query Match 6.9%; Score 104.5; DB 2; Length 1964;
Best Local Similarity 19.8%; Pred. No. 93;
Matches 63; Conservative 47; Mismatches 103; Indels 105; Gaps 13;

Qy 7 EQTEVVKSAIE---TADGALDFNFKYLDQVIPWKTDETIKE-----LSRPFQOYS 55
Db 1166 EQETHLTKLTLEDEARHTEGQIQIRQKHSQAV--EELSELEQTKRLKGNLEKAKQALE 1223
Qy 56 QEASVLVGDIVKLLMSQDKYFEATQTVYEWCGVVTOLLSAYILLFDEYNEKASAKQ-KD 114
Db 1224 GERELANEVTKLQKGD-----SEHKKKVEAQLQE 1256
Qy 115 ILIRI-----LDDGVNKLN---BAQKSLGSSQSFNNASGKLLALDSQLTDFSEK 162
Db 1257 LQVKVTEGDRVRSSELSEKANRLQVELDNVNSLSQSSKSIKLGK-----DFSTL 1306
Qy 163 SSYFQSQVDRIKRAYAGAAAGIVAGPFLIISYIAAGVIEGK---LIPELNDRKAVQ 219
Db 1307 ESQFQDAQELLQEETROK-----LSFSTKLQMEDEKNGLLLEQLEEEEAOK 1353
Qy 220 NFPTSLV-----TV-----KOANKDIDAALKLATAIAIGETKT 255
Db 1354 NLCKQISTLOSQMTDMKKMDENVGSLETVEELKKLQKLEAVNQRFEEKAAAYD--KL 1411
Qy 256 ETETTRFYVDYDDLMLSL 273
Db 1412 EKTTRLQQLDLDISVDL 1429

RESULT 38
T03754
hypothetical protein E - slime mold (Physarum polycephalum) mitochondrion integrated plas
C:Species: mitochondrion Physarum polycephalum
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T03754
R:Nakagawa, C.C.; Jones, E.P.; Miller, D.L.
Curr. Genet. 33, 178-187, 1998
A:Title: Mitochondrial DNA rearrangements associated with mf plasmid integration and plas
A:Reference number: 215055; MUID:98177147; PMID:9508792
A:Accession: T03754
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-640 <NAK>
A:Cross-references: UNIPROT:Q35593; EMBL:AF012249; NID:g2435419; PIDN:AAC15934.1; PID:g24
C:Genetics:
A:Genome: mitochondrion integrated plasmid mf
C:Keywords: mitochondrion

Query Match 6.9%; Score 104; DB 2; Length 640;
Best Local Similarity 22.5%; Pred. No. 23;
Matches 61; Conservative 40; Mismatches 104; Indels 66; Gaps 11;

Qy 10 VEVVKSAIETADGALDFNFKYLDQVIPWKTDETIKELSRPFQOYSQBASVLVGDIVKLL 69
Db 182 LDIMSYASRTAD---QLYNNYIQSLVNLTLPTT-----GVEKVT 219
Qy 70 MDSQDKYFEATQTVYEWCGVVTOLLSAYILLFDEYNEKASAKQDILIRILDDG----- 123
Db 220 KSS-----LSEKVIKMLGYA--LLAGAVAFIGWTKTYQNVLLKADAGSEIVDNGKRITQV 271
Qy 124 VNKLNEAOKSLGSSQSFNNASGKLLALDSQ-----LTNDSEKSSYFQSQVDRI 173
Db 272 VHGVSETDTLLSATNKAVGVAEKLDDVSTKVDTLGVKHAATLTIVNNNTVLFQTLDRV 331
Qy 174 RKEAYAGAAAGIVAGPFLIISYIAAGVIEGKLIPELNDRLKAVQNPFTSLSVTVKQAN 233
Db 332 R--GVLEIQINIDAEPLP-----QQAAILGQNVPEL---LGAINHQIESLAVGLKTN 379
Qy 234 KDIDAANKLKATEIA-----AIGEIKTETTT 260
Db 380 NDLS-----KLANSILNLENQKEIAAQRVT 406

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Authors: Schleich, S.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama, A.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.; Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: E69730
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2334 <KUN>
A;Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469
A;Experimental source: strain 168
R;Toshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A;Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *spo* genes
A;Reference number: Z24350; MUID:95219088; PMID:7704263
A;Accession: T47101
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2334 <YOS>
A;Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782
A;Experimental source: strain BGSC1A1
C;Genetics:
A;Gene: wpa; N17G
C;Superfamily: cell wall-associated protein wpa

Query Match 6.9%; Score 104; DB 2; Length 2334;
Best Local Similarity 25.6%; Pred. No. 1.3e+02;
Matches 60; Conservative 39; Mismatches 85; Indels 50; Gaps 13;

Qy 78 EATQTV-YEWCVVVTVLLSAYILLFPEYNEKK-ASAQKILLIRILD---DGVNKLNEAOK 132
Db 1203 KATESYQYKDGKGVTSVKDAYGTETTYEYKNNNDVTYKQKTEGNTVIAVDGLDVSSETDQ 1262
Qy 133 SLIGSS---QSNN--ASGKLALDSQLTNDPS---EKSSY-FQSQVDRIKREAYAGAA 182
Db 1263 SGKSSAAVYDYGKQIQSSKDLASTNLIKGSFPAQSGWNLTKSKDR-RKISVIADK 1321
Qy 183 AGIVAGPFGI-IISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTVKQAKNDIAAKL 241
Db 1322 SGVLGSKALEVLQSSTAGTDHG-----YSSATQTV-----EL 1355
Qy 242 KLATEIAAIGKIKTETTTFTFYDYDLDLMLLKGAKKMTTC-NEYQORHGK 294
Db 1356 EPNTYTLGKIKTDLAKSRAYFNID-----LRDQDKQIOWHNEYSALAGK 1403

RESULT 41
G97236
ATPase involved in DNA repair [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: G97236
R;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97236
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1163 <KUR>
A;Cross-references: UNIPROT:Q97FK1; GB:AE001437; PIDN:AAK0682.1; PID:gl5025772; GSPDB:G1437
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2736

Query Match 6.9%; Score 103.5; DB 2; Length 1163;
Best Local Similarity 21.7%; Pred. No. 54;
Matches 81; Conservative 55; Mismatches 142; Indels 95; Gaps 14;

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Authors: Schleich, S.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama, A.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.; Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: E69730
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2334 <KUN>
A;Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469
A;Experimental source: strain 168
R;Toshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A;Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *spo* genes
A;Reference number: Z24350; MUID:95219088; PMID:7704263
A;Accession: T47101
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2334 <YOS>
A;Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782
A;Experimental source: strain BGSC1A1
C;Genetics:
A;Gene: wpa; N17G
C;Superfamily: cell wall-associated protein wpa

Query Match 6.9%; Score 104; DB 2; Length 1496;
Best Local Similarity 19.0%; Pred. No. 70;
Matches 62; Conservative 67; Mismatches 124; Indels 74; Gaps 13;

Qy 4 IFAEQTVYEVVKSIAETADGA--LDFNKKYLDQVWPWT-----FDETIKLSRP-----KQ 52
Db 153 IEAEKTVKGMKEMGRDVRVWMEBEKSQVEEKLKWKKEQFKHLEAYEKLKNUFKDSKK 212
Qy 53 EYSQESALVGDIKVL--LMSQDKYFEATQTVYEMC-GVVTQLLSAYILLFDEYNEKKA 109
Db 213 EWEBSKLDDEYSYQTLKDSVTRISEDLQKLLQNCALQ-----EETRRKHL 263
Qy 110 SAQKDLIRILDGVNKLNEAOKSLGSSQSFNNASGKLALDSQLTNDPSFKSYFO-- 167
Db 264 EIQVSEFAKVEDAFACQDARTOL-----DLAGKRDWEVAELRQTLKMDAYFKEM 316
Qy 168 ----SQVDRIKRAYAGAAAGVAGPFGILLISVIAAGVIEG--KLIPELNDRLKAVQN 220
Db 317 KYENGKLEQENRELLGSLK-----ELQENIQSGNSALSGLKKNKFNLEN 362
Qy 221 FFTSLSVT-----VKQANKDIDAAKLKATEIAAIGIKTETE-----TTRFV 264
Db 363 IHKNCANLSKAEWSQVEKVEEINDYKQLQSKAEALKEVELENCRSSTAKMRL 422
Qy 265 DYDDLMLSLKGAAGKMTNCNEYQOR 291
Db 423 QYEBISIMFL-----VLSRTVSEAQR 444

RESULT 40
S32920
cell wall-associated protein precursor wpa [similarity] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: S32920; E69730; T47101
R;Foster, S.J.
Mol. Microbiol. 8, 299-310, 1993
A;Title: Molecular analysis of three major wall-associated proteins of *Bacillus subtilis* protein.
A;Reference number: S32919; MUID:93302506; PMID:8316082
A;Accession: S32920
A;Molecule type: DNA
A;Residues: 1-2334 <FOS>
A;Cross-references: UNIPROT:Q07833; GB:L05634; NID:g304177; PIDN:AAA22883.1; PID:g304179
R;Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea, C.; Bron, S.; Brouillet, S.; Bruchet, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallenbach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel


```
Qy 1 MTGIFABOTVVEVKSIAETADG-----ALDFYNYKLDQVDPKWTDETITK-----ELS 48
Db 112 ITGDEVELEEGAKSVNEKQSEIIGLSLDDETRTV--VLPOQKFSEFLKLEGKERNMLE 169
Qy 49 RF--KQESYQAS-----VLGDIKVLMLDSODKYFEATQTVYEWCGVVT 91
Db 170 RLFNLOEYVGBELSKLARKIRKREKENVLGELGYENINEDVLKERRELLKENNDFN 229
Qy 92 QLLSAYIILPDEYNEKKAQAQKIDILI-----RILDDGVNKLNE-----AQKSLGSSQS--- 140
Db 230 EASKEYLKABEYNEGKEVWGLOIEBEKNRVRKOLMEKKDEIDLKERRARLGSSSKVK 289
Qy 141 -----FNWASKLLALDSQLTN-----DFSEKSSYFQSDVRIRKAYAGAA 182
Db 290 PYIDNYENTLQIDILKEQILSRENTMKAISLEKEDMEKLSIAKDNNKEK----- 339
Qy 183 AGIVAGPFGLLIISYIAAGVTEGKLIPELNDRLKAVQNFPTSLSV-----TVKQANKD 235
Db 340 -----ALPFWIMKHILDAIKKEDKLDLNIKLEKKRLOQKIEKLSLEASNKBELIKONIKD 395
Qy 236 IDAAKLKLATBIAAIGEIKTETTRFRFYVDYDDLMLSLKGAKKMIN-----TCNEYQ 289
Db 396 IDSLTLKIQNLKESKIDNLKVPBE---YKNKINEGIFLLRNYDEKLLKHKHKLGLDCDKFQ 451
Qy 290 ----ORHGKTYLL 298
Db 452 VDFEKAISKKEML 464

RESULT 42
t28667
dynein beta heavy chain - Paramesicium tetraurelia
C:Species: Paramesicium tetraurelia
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28667; PC4340
R:Kandl, K.A.; Forney, J.D.; Asai, D.J.
submitted to the EMBL Data Library, January 1995
A:Description: The dynein genes of Paramesicium: the differential expression of axonemal a
A:Reference number: Z20502
A:Accession: T28667
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4588 <KAN>
A:Cross-references: UNIPROT:Q94709; EMBL:U19464; NID:G625089; PID:G625090; PIDN:AAA61680
R:Asai, D.J.; Beckwith, S.M.; Kandl, K.A.; Keating, H.H.; Tjandra, H.; Forney, J.D.
J. Cell Sci. 107, 839-847, 1994
A:Title: The dynein genes of Paramesicium tetraurelia: Sequences adjacent to the catalytic
A:Reference number: PC4340; MUID:94334383; PMID:8056840
A:Accession: PC4340
A:Molecule type: mRNA
A:Residues: 1886-2085 <ASA>
C:Genetics:
A:Genetic code: SGC5
A:Introns: 43/2; 113/3; 171/3; 4490/1
C:Superfamily: dynein heavy chain, ciliary
C:Keywords: nucleotide binding; P-loop
F:1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match 6.9%; Score 103.5; DB 2; Length 4588;
Best Local Similarity 19.3%; Pred. No. 3.3e+02;
Matches 69; Conservative 67; Mismatches 115; Indels 107; Gaps 15;

Qy 7 BQTVVKSATETADGALDFYNYKLDQVDPKWTDETITKELS-REFKQVYQ--EASVLVGD 52
Db 1207 QETINIKKLDLDFQKVSDFRNDTLNLPYSVHDDMKMDQILYSVTIDEYKLLQNEK 1266
Qy 53 E---YSQASVLVGDIKVLLMDSODKYFEATQTVYEWCGVVTQLLSAYILLF 102
Db 1267 EAADYNQ-----LEKLEFEKSGYKQLRETNVDLKSILKIMWDAISVMNY 1310
Qy 103 EYNEKKA---SAQKDILI---RILDDGVNKLNEAQKSLGSSQSFPNNAASK-LLALDS 153
Db 1311 QYNDWKSFPFQIKADVILLESKNVLGNQLKNLPKEVRNFKGYNALVDKRNMSVVLPLVS 1370
```

```
Qy 154 QLTNDFSEKSSYFQSDVRIRKAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPELND 213
Db 1371 ALHSEFMEDRHW--SQVKDMTKSKF-----SHKAMTFLFD 1403
Qy 214 RLKAVQ--NPFTSLSVTVKQANKIDIDAAKLKLATBIAAIGEI--KTETETTRFYVDYDDL 269
Db 1404 DILALQYLQFDQAQNEVVEVASKEAKTEKLMKTETAWLKQIFEFEDYKETKVFPLDNM 1463
Qy 270 MLSL-----LKGAACKM---INTCNEYQQRHGK-----KTLLEV 300
Db 1464 MEMLDQHSLDLGMNGKGGKYVEFFYNIVEDWREKLGKRVDSVVGWELVKVQKWKTLVNI 1521

RESULT 43
A36014
myosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)
N:Contains: myosin ATPase (BC 3.6.4.1)
C:Species: Drosophila melanogaster
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A36014; B36014
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: cor
A:Reference number: A36014; MUID:90349606; PMID:2117279
A:Accession: A36014
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2017 <KET>
A:Cross-references: UNIPROT:Q99323; GB:M35012
C:Genetics:
A:Gene: FlyBase:zip
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:1-2017/Product: myosin heavy chain, form I #status predicted <MA>
F:46-2017/Product: myosin heavy chain, form II #status predicted <MA2>
F:135-815/Domain: myosin motor domain homology <MMOT>
F:225-232/Region: nucleotide-binding motif A (P-loop)
F:598-631/Region: actin binding #status predicted
F:705-727/Region: actin binding #status predicted
F:888-2017/Domain: coiled coil #status predicted <COI>
F:888-1328/Region: S2
F:1329-2017/Region: light meromyosin
F:176/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:231/Binding site: ATP (Lys) #status predicted
F:745,755/Active site: Cys #status predicted

Query Match 6.8%; Score 103; DB 1; Length 2017;
Best Local Similarity 23.8%; Pred. No. 1.2e+02;
Matches 65; Conservative 44; Mismatches 102; Indels 62; Gaps 13;

Qy 8 QTVVEVKSATETADGALDFYNYKLDQVDPKWTDETITKELS-REFKQVYQ--EASVLVGD 64
Db 1357 ENLRKAKTVLEKAGTLEAEN--ADLATELRSVNSRQENDRRRRKQAEQAELQVLUKE 1314
Qy 65 IKVLLMDSODKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILRILDDGV 124
Db 1315 IERARSELQEK---CTKLQQAENITNQL-----EEAELKASA----AVKSASNME 1358
Qy 125 NKLENAQKSL-----LGSSQSPFNNAAGKLLALDSQLTNDPSEKSSY-----FQSQVD 171
Db 1359 SQLTEAQQLLEETPRQKLGSLSKLRQIESEKEALQELEDDEAKRNYERKLAEVTTQM 1418
Qy 172 RIRKEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQ 231
Db 1419 EIRKKAEDA-----DLAKELEEK--KRLNKDIEALER-----QVKE 1454
Qy 232 ---ANKDIDAAKLKLATBIAAIGEIKTETTR 261
Db 1455 LIAQNDRLDKSKKKIQSELED-ATIELEAQRTK 1486
```

Wed Feb 2 11:26:41 2005

RESULT 44

S61477
myosin II heavy chain, non-muscle - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C:Accession: S61477; S65349
R:Manuscript: S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
J. Mol. Biol. 255, 98-109, 1996
A:Title: Molecular organization and alternative splicing in zipper, the gene that encodes the myosin heavy chain; myosin heavy chain; myosin motor domain homology
A:Reference number: S61477; MUID:96144835; PMID:8568878
A:Accession: S61477
A:Molecule type: DNA
A:Residues: 1-2057 <MAN>
A:Cross-references: UNIPROT:Q94987; EMBL:U35816
R:Manuscript: S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
submitted to the EMBL Data Library, September 1995
A:Reference number: S65349
A:Accession: S65349
A:Molecule type: DNA
A:Residues: 1-1908, 'NL', 1911-2057 <MAN>
A:Cross-references: EMBL:U35816; NID:g1141789; PIDN:AA09049.1; PID:g1572481
C:Genetics:
A:Gene: zip
A:Cross-references: FlyBase:FBgn0005634
C:Introns: 10/3; 90/3; 162/3; 215/1; 255/3; 263/3; 303/3; 334/3; 1248/1; 1349/3; 1526/3;
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: alternative splicing; ATP; nucleotide binding; P-loop
F:135-855/Domain: myosin motor domain homology <NMOT>
F:225-232/Region: nucleotide-binding motif A (P-loop)

Query Match 6.8%; Score 103; DB 2; Length 2057;

Best Local Similarity 23.8%; Pred. No. 1.2e+02; Indels 62; Gaps 13;
Matches 65; Conservative 44; Mismatches 102;

QY 8 QTVEVVKSAETADGALDFYKYLQVDPKTFDEIKELSRFKQESYQ--EASVLVGD 64
DB 1297 ENLRKAKTVLEKAKGLEAEN--ADLATELSVNSRRQENRRRKOAESQIAELQVKLAE 1354
QY 65 IKVLLMDSQDKYFEATQTVYVCGVVTOLLISAVILLDFEYNEKKASAKDILIRILDGV 124
DB 1355 IERARSELQEK--CTKLOQEAENITNOL-----EEALIKASA-----AVKASWNE 1398
QY 125 NKLNEAKSL-----LGSSQSFNNASGKLLALDSQLTNDFFSEKSY-----FQSQVD 171
DB 1399 SQTEAQQLEETRQKLGSLSKLRQIESEKALQQLDEEDDEAKNYERKLAETVTQMQ 1458
QY 172 RIRKEAYAGAAAGIVAGPFGLLIISYSIAAGVIRGKLIPELNDRLKAVQNFFTSLSVTKQ 231
DB 1459 EIKKKAEDA-----DLAKELEEGK--KRLNKDIEALER-----QVKE 1494
QY 232 ---ANKDIDAKKLATEIAAIGETETETTR 261
DB 1495 LIAQNDRLDKSKKIQSELED-ATIELEAORTK 1526

RESULT 45

AD2742
conserved hypothetical protein Atul348 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: *Agrobacterium tumefaciens*
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AD2742
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD2742
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2155 <KUR>

A:Cross-references: UNIPROT:Q8UFP9; GB:AB008688; PIDN:AA42354.1; PID:g17739760; GSPDB:GT
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul348
A:Map position: circular chromosome

Query Match 6.8%; Score 103; DB 2; Length 2155;
Best Local Similarity 20.2%; Pred. No. 1.3e+02;
Matches 71; Conservative 64; Mismatches 116; Indels 100; Gaps 13;

QY 3 GIFAPQTVVVKSAI-----ETADGALDFYKYLQVDPKTFDEIKELSRFKQESYQEA 58
DB 1428 GVISE-TLTSVESALDARGESIRSDNRTRELNSMLASRS-----AELSLIEE---KA 1478
QY 59 SVLVGDIKVLMLDSODKYFEATQTVYE-----WCGVVTQLLSAYILLDFEYNEKKAS 110
DB 1479 KPVEEYATIGREAAEKIVSAAQSAELLSQNSGMVGMVEQAISDYAAAGTDAASKLVA 1538
QY 111 AQKDILIRILDGVNKLNEAOKSLGSSQSFNNASGKLLALDSQLTNDFFS----- 160
DB 1539 ATRQSTDMLSETHTAMADAVEQIDKYARSGSDAANKLVAATROSTDMLSQTHNSMAEIV 1598
QY 161 -EKSSYFQSQVDRIKAEYAGAAAGIVAGPFGLLIS-----YSIAAGVIEG 205
DB 1599 EQGATNFNAVER-----AAQFGGADEALNASATRFSESASQAADMVSSSRLLLEG 1650
QY 206 KLIPELNDRLK-----AVQNFSTSLSVTKQANKOIDAAKLAETIAAIGSIK 254
DB 1651 KI-----DRLSNISGQTLAQVAGIVGRPEHSHKVLQSAELNAAQSSD-----VGTLE 1699
QY 255 TETETRFYVDYDMLMLSLKGAKK-----MINTCNEYOQR 291
DB 1700 -----ERQDALRSLSVGLVKRSEIEIETAMRVGVVENTLNEAER 1740

RESULT 46

C97523
hypothetical protein AGR_C_2490 [imported] - Agrobacterium tumefaciens (strain C58, Cerec)
C:Species: *Agrobacterium tumefaciens*
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: C97523
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, P.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: C97523
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2155 <KUR>
A:Cross-references: UNIPROT:Q8UFP9; GB:AB007869; PIDN:AAK87140.1; PID:g15156408; GSPDB:GT
C:Genetics:
A:Gene: AGR_C_2490
A:Map position: circular chromosome

Query Match 6.8%; Score 103; DB 2; Length 2155;
Best Local Similarity 20.2%; Pred. No. 1.3e+02;
Matches 71; Conservative 64; Mismatches 116; Indels 100; Gaps 13;

QY 3 GIFAPQTVVVKSAI-----ETADGALDFYKYLQVDPKTFDEIKELSRFKQESYQEA 58
DB 1428 GVISE-TLTSVESALDARGESIRSDNRTRELNSMLASRS-----AELSLIEE---KA 1478
QY 59 SVLVGDIKVLMLDSODKYFEATQTVYE-----WCGVVTQLLSAYILLDFEYNEKKAS 110
DB 1479 KPVEEYATIGREAAEKIVSAAQSAELLSQNSGMVGMVEQAISDYAAAGTDAASKLVA 1538
QY 111 AQKDILIRILDGVNKLNEAOKSLGSSQSFNNASGKLLALDSQLTNDFFS----- 160
DB 1539 ATRQSTDMLSETHTAMADAVEQIDKYARSGSDAANKLVAATROSTDMLSQTHNSMAEIV 1598
QY 161 -EKSSYFQSQVDRIKAEYAGAAAGIVAGPFGLLIS-----YSIAAGVIEG 205

Db 1599 EOSATFNFAVER-----AAQFGAADALNASATRFSESQAQADVMSSSRLLLEG 1650

Qy 206 KLIPBLNDRK-----AVQFFTSLSVTVKQANKDIDAAKLKLATEIAAIGBIK 254

Db 1651 KI-----DRLSNISQTLAQAGIVGRFEHSKVLQSASELLNAAQSSL-----VGTLE 1699

Qy 255 TETETTRYVDYDDMLSLKGAANK-----MINTONEVQOR 291

Db 1700 -----BRQDALRSLVGLVKRSBEIETAMENVVGVVENTLNEAER 1740

RESULT 47

D97198

methyl-accepting chemotaxis protein [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C:Accession: D97198

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D97198

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-379 <KUR>

A:Cross-references: UNIPROT:Q97GE7; GB:AE001437; PIDN:AAK80375.1; PID:gl5025436; GSPDB:CAC2420

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2420

Query Match 6.8%; Score 102.5; DB 2; Length 379;

Best Local Similarity 21.2%; Pred. No. 14;

Matches 58; Conservative 40; Mismatches 83; Indels 93; Gaps 13;

Qy 72 SODKYFEATQTYEWCVVTTQLLSAYILLFDEYNEK-----KASAKQDI-- 115

Db 51 SNDEFGNLTSHFF-----TMSQNLKNLIIQIKESSRLEDVSTEMLKSCSESASASSDISQ 106

Qy 116 -LIRILD--GVNKLNEAQSLLG-----SSQSFNNAAGK----- 147

Db 107 NIVNVADSNVSLNKLNANDSVMLIDVFSQVILNVNIICTKVEETSYVSKGKMDINK 166

Qy 148 ----LLALDSQTLNDFSEKSSYFQSOVDRIKAYAGAAAGIV-----A 187

Db 167 ITEQMLNIENRV-NSLSEVISKLTCHDINN--YSNVISDIABETNLSLNANIEAARA 223

Qy 188 GPFGLIISYSTAAGVIE-----GKLIPELN-DRLKAVQNF-----FTSLSV 227

Db 224 GENGK--GFSVVADKIRSLADESGKAASEIGNLIREMNEISIKAINSMNAGKKEVTSGKV 281

Qy 228 TVQKANKDIDAAKLKLATEIAAIGBIKTETETTR 261

Db 282 IVEQNLTFNEIANSINSILAATSEVTASMTVK 315

RESULT 48

H81665

replicative DNA helicase TC0784 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004

C:Accession: H81665

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: H81665

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-472 <TET>

A:Cross-references: UNIPROT:Q9PJP4; GB:AE002346; GB:AE002160; NID:g7190805; PIDN:AAF3958

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0784

C:Superfamily: phage P22 gene 12 protein

Query Match 6.8%; Score 102.5; DB 2; Length 472;

Best Local Similarity 23.1%; Pred. No. 19;

Matches 74; Conservative 42; Mismatches 105; Indels 99; Gaps 17;

Qy 25 DFYKNYLDQVIPWKTFFDETIK-----ELSRFKQYSQSEASVLLVGDIDIKVLLMDS 72

Db 48 DFY--FLEHRIIFRVLDQAFKSDRPMPLHTGELKR-----RDQLNVIGASVLI--- 96

Qy 73 QDKYFEATQTYEWCVVTTQLLSAYILLFDEYNE-----KASAKQDILIRILD-- 121

Db 97 -----TLSEFAGT-----SAYI---EEYADIIRSKSILRKMIQAADIEKKAEEP 139

Qy 122 -DGVNKLNEAQSLLGSSQSFNNA-----SGKLALDSQTLNDF-----SEKSSYFOSQVD 171

Db 140 RDVTALDDAQNLLFRISQITTFAPYVVLVSDKLGSLSTKDKSFLALQERQEAFOASSH 199

Qy 172 RIRKEAYAGAAAGIV-----AGPFGLIISYSIAAGVIEGKLIPELNDRLKAVQNF-- 221

Db 200 DARIPTLSGGPTHFLDLDRMLNGFSPNLI---LAARPAMGKTALALN-----IVENFCF 252

Qy 222 -----FTSLSVTVKQ-----ANKDIDAAKLKLATEIA-----AIGIKTETETTR 261

Db 253 ESRLPVGIFSLMTVTDQILHRIICRSSEVEAKISVG-DISGRDFQRVVSVVREMEHTL 311

Qy 262 FYVDYDDMLSLKGAAKKM 281

Db 312 LIDDYPLGKITDLRARARM 331

RESULT 49

T44825

hypochemical protein wzc [imported] - Acinetobacter lwoffii

N:Alternate names: protein tyrosine kinase

C:Species: Acinetobacter lwoffii

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004

C:Accession: T44825

R:Nakar, D.; Gutnick, D.L.

submitted to the EMBL Data Library, July 1999

A:Description: Genomic organization of the wzc region of Acinetobacter lwoffii RAG-1 req

A:Reference number: Z22856

A:Accession: T44825

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-726 <NAK>

A:Cross-references: UNIPROT:Q9RMD9; EMBL:AJ243431; PIDN:CAB57193.1

A:Experimental source: strain RAG-1

C:Genetics:

A:Gene: wzc

Query Match 6.8%; Score 102; DB 2; Length 726;

Best Local Similarity 21.7%; Pred. No. 36;

Matches 67; Conservative 48; Mismatches 92; Indels 102; Gaps 13;

Qy 8 QTVBWKSAIETADGALDFYKNYLDQVIPWKT--FDETIKLSRFKQYSQSEASV----- 60

Db 173 KTDEIVLTA-----PLNKLNLQKNQFGTWKVAIFTDNTPDANVFKNSLPAAVQAIL 226

Qy 61 -----LVGDIKVLMDSDQKYFEATQTYEWCVVTTQLLSAYILLFDEYNEKKA 111

Db 227 NYSVAERGLTGT-ILGLNYQODKEH-----ITKVNAILAAYSAQNVERRSA 273

Qy 112 QKDTLIRILDGV-----NKLNEAQSLLGSSQSFNNAAGKLLALDSQTLNDFSEKSSYFQ 167

Db 274 ETAQTLFLDEQLPDLKQLDDAERVFNFKQQYN-----TVDVTKESELYL 320

Qy 168 SOVDRIR-----KEAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIPELNDRLKAVQNF 221

Db 321 TQSTLTETTKIELOQQOQADLAUKYTA-----EHPAIREINAQITALDKQ 364

Wed Feb 2 11:26:41 2005

QY 222 FTSLSVTVKQANKDIDAALKLATEIAAIGAIKIKETETTRFYVDYDDLMLSLKGAAKM 281
Db 365 ITQLNSTLUQL-PDIQRYLQLFREV-----EVKIQ-----LYTALL----- 400

QY 282 INTCNEYQQ 290
Db 401 ----NSYQQ 405

RESULT 50
H75001
methyl-accepting chemotaxis protein (tlpc) PAB1336 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: H75001
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: H75001
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-739 <RAW>
A:Cross-references: UNIPROT:Q9UYF8; GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB5045
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1336

Query Match 6.8%; Score 102; DB 2; Length 739;
Best Local Similarity 19.5%; Pred. No. 37;
Matches 64; Conservative 73; Mismatches 140; Indels 52; Gaps 11;
QY 1 MTGIFAEQTVVVKSA--IETADGALDFYKVKLDQVIPKWTDFETIKELSRFKQEYSQEA 58
Db 425 VTETRESIGSLVEMANDLEKRANALAQVSKDVTEAI--NQVNEAIQQVSIQAQOQETI 482
QY 59 SVLVGDIKVLMDSQDKYFEATQTVVCGVTVTOLLISAVILLDFEYNEKKAQAQDI--- 115
Db 483 NEITDGMRLVAQTS-----ESVRAMEEFGNAVTEVVS-----IANEGSQKGDALAKRIEDI 534
QY 116 --LIRIILDDGVNKLNE-----AQKSLIGSSOSFNNA-----SGKLLALDS 153
Db 535 QHMSRIEETVSKVAEMSRNIEITNVTISIAEQTNLLALNAAIEAARAGEAGRGFAVVA 594
QY 154 QLTNDFSEKSY-----FQSQVDIR--KEAYAGAAGI-VAGPFGLIISYSIAAGVIRG 205
Db 595 QEIRKLAEESQKQADNIKSIIIDKIDKEAVEATKEGVSVIGSESSEITLRTIIGYLANIA 654
QY 206 KLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAALKLATEIAAIGAIKIKETETTRFYVD 265
Db 655 TLLQETSERM-----TTVKEQIVRTQEEVDKRALLENLAASABETTASAEVSSAIE 707
QY 266 YDDLMLSLKGAAKKMINTCNEYQORHGK 294
Db 708 QOTAAIEELRRAQELKDMVGRMEQIVGK 736

Search completed: January 28, 2005, 19:16:48
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:06:39 ; Search time 198 Seconds
(without alignments)
880.498 Million cell updates/sec

Title: US-09-993-292B-24
Perfect score: 1510
Sequence: 1 MTGIFAEQTVVKSIAETA.....TCNEYQQRHGKTLLEVPDI 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database : UniProt 02.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1505	99.7	302	1	HLVE_SALPA	Q93rr6 salmonella
2	1467	97.2	302	1	HLVE_SALTI	Q8z727 salmonella
3	1376	91.1	302	1	HLVE_ECOLI	P77335 escherichia
4	1361	90.1	302	1	HLVE_ECO57	Q9reb3 escherichia
5	1104	73.1	300	2	Q9X2S8	Q9x2s8 escherichia
6	513	34.0	113	1	HLBL_SHIFL	Q9rct3 shigella fl
7	382	25.3	93	1	HLBL_ECOL6	Q8fi27 escherichia
8	136.5	9.0	895	2	Q9LIW7	Q9liw7 oryza sativ
9	132	8.7	495	2	P71497	P71497 mycoplasma
10	130	8.6	2723	2	Q7RQB6	Q7rgb6 mycoplasma
11	128.5	8.5	713	2	Q6MT03	Q6mt03 mycoplasma
12	128.5	8.5	713	2	CAE77235	CAE77235 mycoplasma
13	126.5	8.4	577	2	Q7N5U5	Q7n5u5 photorhabd
14	125	8.3	1521	2	Q6G015	Q6g015 bartonella
15	124.5	8.2	1825	2	Q7RIT2	Q7rit2 plasmodium
16	122.5	8.1	4007	2	Q7SHZ4	Q7shz4 neurospora
17	121	8.0	2664	2	Q7REL0	Q7rel0 plasmodium
18	120.5	8.0	693	2	Q7N9I8	Q7n9i8 wolinnella s
19	117.5	7.8	555	2	Q7VI37	Q7vi37 helicobacte
20	117	7.7	2139	2	Q07569	Q07569 entamoeba h
21	116.5	7.7	478	2	Q64067	Q64067 bacterioph
22	116.5	7.7	478	2	Q31954	Q31954 bacillus su
23	116.5	7.7	652	2	Q7CMF0	Q7cmf0 bacillus an
24	116.5	7.7	652	2	Q9X360	Q9x360 bacillus an
25	116.5	7.7	652	2	AAT28865	AAT28865 bacillus
26	116.5	7.7	779	2	Q7RM79	Q7rm79 plasmodium
27	116	7.7	541	2	Q7ADE2	Q7ade2 geobacter s
28	116	7.7	541	2	AAR34750	AAR34750 geobacter
29	116	7.7	941	2	Q81Z98	Q81z98 bacillus an
30	116	7.7	941	2	AAT29469	AAT29469 bacillus
31	116	7.7	1023	1	HLVI_ECOLI	P09983 escherichia

32	115	7.6	1005	1	RA50_METJA	Q58718 methanococc
33	115	7.6	1093	2	Q288G7	Q288g7 listeria in
34	114.5	7.6	821	2	Q08581	Q08581 saccharomyc
35	113.5	7.5	587	2	Q97WH8	Q97wh8 sulfolobus
36	113.5	7.5	1478	2	Q7RDZ0	Q7rdz0 plasmodium
37	112.5	7.5	520	1	REC_N_AQUAE	Q68834 aquifex aeo
38	112.5	7.5	584	1	SYR_SINY3	Q55486 synechocyst
39	112.5	7.5	941	2	Q6CGC3	Q6cgc3 yarrowia li
40	112.5	7.5	1024	1	HLVA_ECOLI	P08715 escherichia
41	112.5	7.5	1098	2	Q50733	Q50733 borrelia bu
42	112.5	7.5	1906	2	Q6C359	Q6c359 yarrowia li
43	112.5	7.5	3102	2	Q45614	Q45614 caenorhabdi
44	112	7.4	1012	2	Q6GNE7	Q6gne7 xenopus lae
45	112	7.4	1081	2	Q7RFQ7	Q7rfq7 plasmodium
46	112	7.4	1464	2	Q81IF6	Q81if6 plasmodium
47	111	7.4	329	1	ESG2_TRYBB	P04478 trypanosoma
48	111	7.4	329	2	Q6ZM87	Q6zm87 trypanosoma
49	111	7.4	329	2	CAD21890	CAD21890 trypanoso
50	111	7.4	465	2	Q73AJ3	Q73aj3 bacillus ce
51	111	7.4	465	2	AAS40708	AAS40708 bacillus
52	111	7.4	1033	2	Q739A1	Q739a1 bacillus ce
53	111	7.4	1033	2	AAS41161	AAS41161 bacillus
54	111	7.4	1189	2	Q8CP10	Q8cp10 staphylococ
55	111	7.4	1295	2	Q22257	Q22257 caenorhabdi
56	111	7.4	2033	2	Q7XEH4	Q7xeh4 oryza sativ
57	110.5	7.3	1489	2	Q7RJ32	Q7rj32 plasmodium
58	110.5	7.3	2867	1	RBP2_PLAVB	Q00799 plasmodium
59	110	7.3	411	2	Q8D3E6	Q8d3e6 wiggleswort
60	110	7.3	957	2	Q7RNH1	Q7rnh1 plasmodium
61	110	7.3	1285	2	Q9XUJ3	Q9xuj3 thermotoga
62	109.5	7.3	504	2	Q81752	Q81752 plasmodium
63	109.5	7.3	1024	2	Q8G9Z4	Q8g9z4 escherichia
64	109.5	7.3	1024	2	Q8GA40	Q8ga40 escherichia
65	109.5	7.3	1024	2	Q8FE01	Q8fe01 escherichia
66	109.5	7.3	2749	2	Q7REY3	Q7rey3 plasmodium
67	109	7.2	541	2	Q7X8S6	Q7x8s6 oryza sativ
68	109	7.2	541	2	CAE01868	CAE01868 oryza sat
69	109	7.2	573	2	Q97DG6	Q97dg6 clostridium
70	109	7.2	921	2	Q7P2Q9	Q7p2q9 fusobacteri
71	109	7.2	1363	2	Q874Y4	Q874y4 podospora a
72	109	7.2	1492	2	Q68724	Q68724 yersinia pe
73	109	7.2	1545	2	Q9ZH03	Q9zh03 yersinia pe
74	109	7.2	1545	2	Q7ARG4	Q7arg4 yersinia pe
75	109	7.2	1545	2	AAS58642	AAS58642 yersinia
76	108.5	7.2	852	2	Q81754	Q81754 plasmodium
77	108.5	7.2	884	2	Q8PW31	Q8pw31 methanosarc
78	108.5	7.2	953	2	Q73E75	Q73e75 bacillus ce
79	108.5	7.2	953	2	AAS39419	AAS39419 bacillus
80	108.5	7.2	1175	2	Q8XNW6	Q8xnw6 clostridium
81	108.5	7.2	1273	2	Q9BL02	Q9bl02 caenorhabdi
82	108.5	7.2	1959	1	MYH9_CHICK	P14105 gallus gall
83	108.5	7.2	1999	2	Q63731	Q63731 rattus norv
84	108	7.2	622	1	SR68_CAEBL	Q20822 caenorhabdi
85	108	7.2	723	2	Q8F0P9	Q8f0p9 mesoplasma
86	108	7.2	792	2	Q9ARQ9	Q9arq9 oryza sativ
87	108	7.2	872	2	Q7NFE9	Q7nfe9 gloeobacter
88	108	7.2	1877	2	Q6BFD6	Q6bfd6 paramescium
89	108	7.2	2097	2	Q7RZF8	Q7rzf8 neurospora
90	108	7.2	2115	2	Q6MFH6	Q6mfh6 neurospora
91	108	7.2	2115	2	CAE85613	CAE85613 neurospor
92	107.5	7.1	534	2	Q6FNN2	Q6fnn2 homo sapien
93	107.5	7.1	534	2	AAT01278	AAT01278 homo sapi
94	107.5	7.1	720	2	Q9H6Q7	Q9h6q7 homo sapien
95	107	7.1	318	2	Q26749	Q26749 trypanosoma
96	107	7.1	1611	2	Q7RD43	Q7rd43 plasmodium
97	107	7.1	4368	2	Q61851	Q61851 caenorhabdi
98	106.5	7.1	527	2	Q74JAO	Q74jao lactobacill
99	106.5	7.1	527	2	AAS09030	AAS09030 lactobacil
100	106.5	7.1	1938	1	MYSD_CAEBL	P02567 caenorhabdi

ALIGNMENTS

RESULT 1		
HYE SALPA	STANDARD;	PRT; 302 AA.
ID	HYE SALPA	
AC	Q93R66;	
DT	29-MAR-2004 (Rel. 43, Created)	
DT	29-MAR-2004 (Rel. 43, Last sequence update)	
DT	05-JUL-2004 (Rel. 44, Last annotation update)	
DE	Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).	
GN	Name=hlyE; Synonyms=clyA, sheA;	
OS	Salmonella paratyphi-a	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC	Enterobacteriaceae; Salmonella.	
OX	NCBI_TaxID=54388;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SMI S3068/99;	
RX	MEDLINE=22215712; PubMed=12228306;	
RA	Oscarsson J., Westermarck M., Loefdaahl S., Olsen B., Palmgren H.,	
RA	Mizunoe Y., Wai S.N., Uhlin B.E.;	
RT	"Characterization of a pore-forming cytotoxin expressed by Salmonella	
RT	enterica serovars typhi and paratyphi A.";	
RL	Infect. Immun. 70:5759-5769(2002).	
CC	-!- FUNCTION: Toxin, which has some hemolytic activity towards	
CC	mammalian cells. Acts by forming a pore-like structure upon	
CC	contact with mammalian cells (By similarity).	
CC	-!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a	
CC	monomer, while in outer membrane vesicles, it oligomerizes to form	
CC	a pore structure that is active. Probably forms an octamer (By	
CC	similarity).	
CC	-!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer	
CC	membrane vesicles. Also found in the periplasmic space (By	
CC	similarity).	
CC	-!- PTM: In periplasm, it forms a disulfide bond, which prevents the	
CC	oligomerization. In outer membrane vesicles, the redox status	
CC	prevents formation of the disulfide bond, leading to	
CC	oligomerization and pore formation (By similarity).	
CC	-!- SIMILARITY: Belongs to the hemolysin E family.	
CC	-----	
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AJ313033; CAC38362.1; --	
DR	HSSP; P77335; 1QOV.	
DR	InterPro; IPR010356; HlyE.	
DR	Pfam; PF06109; HlyE; 1.	
KW	Cytolysis; Hemolysis; Toxin; Transmembrane.	
FT	INIT MET 0 By similarity.	
FT	TRANSMEM 178 198 Potential.	
FT	DISULFID 86 284 In monomeric form (By similarity).	
SQ	SEQUENCE 302 AA; 33613 MW; 22DB9112B763A061 CRC64;	
Query Match		
Best Local Similarity 100.0%; Pred. No. 1.4e-91;		
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	2	TGIFAEQTVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFKQVSEASVL 61
DB	1	TGIFAEQTVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFKQVSEASVL 60
QY	62	VGDIKVLMDSDQKYFEATQTVYEWCGVTVQLLSAYILLFDEYNEKASAKDILIRILD 121
DB	61	VGDIKVLMDSDQKYFEATQTVYEWCGVTVQLLSAYILLFDEYNEKASAKDILIRILD 120
QY	122	DGVNKLNEAKSLIGSSQFNNSAGKLLDLSQITNDFSEKSSYFQSDVRIRKEAYAGA 181
DB	121	DGVNKLNEAKSLIGSSQFNNSAGKLLDLSQITNDFSEKSSYFQSDVRIRKEAYAGA 180

QY	182	AAGIVAGPRLIISYSIAAGVIEGKLIPELNRLKAVQNFFTSLSVTVKQAKNDIDAAL 241
DB	181	AAGIVAGPRLIISYSIAAGVIEGKLIPELNRLKAVQNFFTSLSVTVKQAKNDIDAAL 240
QY	242	KLATEIAAIGEIKTETETTRFYVDYDMLSLKGAAKMINTCNEYQORHGKKTLLVP 301
DB	241	KLATEIAAIGEIKTETETTRFYVDYDMLSLKGAAKMINTCNEYQORHGKKTLLVP 300
QY	302	DI 303
DB	301	DI 302

RESULT 2		
HYE SALTI	STANDARD;	PRT; 302 AA.
ID	HYE SALTI	
AC	Q8Z727; Q934C4;	
DT	29-MAR-2004 (Rel. 43, Created)	
DT	29-MAR-2004 (Rel. 43, Last sequence update)	
DT	01-OCT-2004 (Rel. 45, Last annotation update)	
DE	Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).	
GN	Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=STV1498, t1477;	
OS	Salmonella typhi.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC	Enterobacteriaceae; Salmonella.	
OX	NCBI_TaxID=601;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Ty21a, and SMI S2369/96;	
RX	MEDLINE=22215712; PubMed=12228306;	
RA	Oscarsson J., Westermarck M., Loefdaahl S., Olsen B., Palmgren H.,	
RA	Mizunoe Y., Wai S.N., Uhlin B.E.;	
RT	"Characterization of a pore-forming cytotoxin expressed by Salmonella	
RT	enterica serovars typhi and paratyphi A.";	
RL	Infect. Immun. 70:5759-5769(2002).	
CC	-!- FUNCTION: Toxin, which has some hemolytic activity towards	
CC	mammalian cells. Acts by forming a pore-like structure upon	
CC	contact with mammalian cells (By similarity).	
CC	-!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a	
CC	monomer, while in outer membrane vesicles, it oligomerizes to form	
CC	a pore structure that is active. Probably forms an octamer (By	
CC	similarity).	
CC	-!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer	
CC	membrane vesicles. Also found in the periplasmic space (By	
CC	similarity).	
CC	-!- PTM: In periplasm, it forms a disulfide bond, which prevents the	
CC	oligomerization. In outer membrane vesicles, the redox status	
CC	prevents formation of the disulfide bond, leading to	
CC	oligomerization and pore formation (By similarity).	
CC	-!- SIMILARITY: Belongs to the hemolysin E family.	
CC	-----	
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AJ313033; CAC38362.1; --	
DR	HSSP; P77335; 1QOV.	
DR	InterPro; IPR010356; HlyE.	
DR	Pfam; PF06109; HlyE; 1.	
KW	Cytolysis; Hemolysis; Toxin; Transmembrane.	
FT	INIT MET 0 By similarity.	
FT	TRANSMEM 178 198 Potential.	
FT	DISULFID 86 284 In monomeric form (By similarity).	
SQ	SEQUENCE 302 AA; 33613 MW; 22DB9112B763A061 CRC64;	
Query Match		
Best Local Similarity 100.0%; Pred. No. 1.4e-91;		
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	2	TGIFAEQTVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFKQVSEASVL 61
DB	1	TGIFAEQTVWVKSATETADGALDFYKYLQVLPKTFDETIKELSRFKQVSEASVL 60
QY	62	VGDIKVLMDSDQKYFEATQTVYEWCGVTVQLLSAYILLFDEYNEKASAKDILIRILD 121
DB	61	VGDIKVLMDSDQKYFEATQTVYEWCGVTVQLLSAYILLFDEYNEKASAKDILIRILD 120
QY	122	DGVNKLNEAKSLIGSSQFNNSAGKLLDLSQITNDFSEKSSYFQSDVRIRKEAYAGA 181
DB	121	DGVNKLNEAKSLIGSSQFNNSAGKLLDLSQITNDFSEKSSYFQSDVRIRKEAYAGA 180

```

CC  -!- SIMILARITY: Belongs to the hemolysin E family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AJ313032; CAC38360.1; -
DR  EMBL; AJ313034; CAC38363.1; -
DR  EMBL; AL627270; CAD01758.1; ALT_INIT.
DR  EMBL; AE016839; AAO69115.1; ALT_INIT.
DR  HSSP; P77335; 100Y.
DR  InterPro; IPR010356; HlyE.
DR  Pfam; PF06109; HlyE; 1
KW  Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
FT  INIT_MET 0 By similarity.
FT  TRANSMEM 178 198 Potential.
FT  DISULFID 86 284 In monomeric form (By similarity).
SQ  SEQUENCE 302 AA; 33658 MW; 2BDFD835D044FDAE CRC64;

Query Match 97.2%; Score 1467; DB 1; Length 302;
Best Local Similarity 97.4%; Pred. No. 4.6e-89;
Matches 294; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TGIFAEQVEVVKSAIETADGALDPYKYLQVLPWKTFDETIKLSRFKQYSEASVL 61
Db 1 TGIFAEQVEVVKSAIETADGALDYLQVLPWKTFDETIKLSRFKQYSEASVL 60

Qy 62 VGDIKVLLMDSQDKYFEATQTVYECGVVTVQLLSAYILLFDEYNEKKASAKDILIRILD 121
Db 61 VGDIKVLLMDSQDKYFEATQTVYECGVVTVQLLSAYILLFDEYNEKKASAKDILIRILD 120

Qy 122 DVNKLNAQKSLGSSQSFNNASGKLLALDSQLTNDSEKSSYFQSDVRIRKEAYAGA 181
Db 121 DGVKLNLAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSDVRIRKEAYAGA 180

Qy 182 AAGIVAGPFGLLIYSIAAGVIEGKLIPELNDRLKAVONFTLSVTVKQAKDIDAANKL 241
Db 181 AAGIVAGPFGLLIYSIAAGVIEGKLIPELNNRLKATVQNFSTLSATVQAKDIDAANKL 240

Qy 242 KLATEIAAIGEIKTETETTRFYVDYDDLMLSLKGAKKMINTCNEYQQRHGKTLLEVP 301
Db 241 KLATEIAAIGEIKTETETTRFYVDYDDLMLSLKGAKKMINTCNEYQQRHGKTLLEVP 300

Qy 302 DI 303
Db 301 DV 302

RESULT 3
HlyE_ECOLI STANDARD; PRT; 302 AA.
AC P77335; Q47276; Q8VU70; Q9R3G4;
DT 01-NOV-1997 (Rel. 35, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hemolysin E, chromosomal (hemolysis-inducing protein) (Silent
DE hemolysin sheA) (Cytotoxin clyA) (latent pore-forming 34 kDa
DE haemolysin).
GN Name=hlyE; Synonyms=clyA, sheA, hpr; OrderedLocusNames=bl182;
OS Escherichia coli
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / XLI-BLUE;
RA McNameara P.J., Iandolo J.J., Uhlich G.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;
RT "The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted
RT haemolysin."
RL Mol. Microbiol. 25:107-115(1997).
RP [3]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-12, SUBCELLULAR LOCATION, AND
RP INDUCTION.
RC STRAIN=K12;
EX MEDLINE=99157562; PubMed=10027972;
RA Ludwig A., Bauer S., Benz R., Bergmann B., Goebel W.;
RT "Analysis of the SlyA-controlled expression, subcellular localization
RT and pore-forming activity of a 34 kDa haemolysin (ClyA) from
RT Escherichia coli K-12."
RL Mol. Microbiol. 31:557-567(1999).
RP [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
EX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RP [6]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
EX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RP [7]
RP SEQUENCE FROM N.A.
RC STRAIN=CH9802;
RA Chang G.-N., Ho K.-C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RP [8]
RP SEQUENCE OF 1-295 FROM N.A.
RC STRAIN=K12 / XLI-BLUE;
RA King C.H., Shinnick T.M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RP [9]
RP SEQUENCE OF 1-155 FROM N.A.
RC STRAIN=K12 / AB1157;
RA Woodgate R.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RP [10]
RP SEQUENCE OF 1-17, AND MUTAGENESIS OF 87-GLY--VAL-89; 142-ASN-ALA-143;
RP 182-ALA-GLY-183; 186-ALA-GLY-187; ASP-267 AND 292-GLY-LYS-293.
RX MEDLINE=99316011; PubMed=10383763;
RA Oscarson J., Mizunoe Y., Li L., Lai X.-H., Wieslander A., Uhlin B.E.;
RT "Molecular analysis of the cytolytic protein ClyA (SheA) from
RT Escherichia coli."
RL Mol. Microbiol. 32:1226-1238(1999).
RP [11]
RP PARTIAL SEQUENCE, FUNCTION, SUBCELLULAR LOCATION, OLIGOMERIZATION, AND
RP DISULFIDE BOND FORMATION.
EX MEDLINE=22894281; PubMed=14532000;
RA Wal S.N., Lindmark B., Soederblom T., Takade A., Westermark M.,
RA Oscarsson J., Jass J., Richter-Dahlfors A., Mizunoe Y., Uhlin B.E.;

```

"Vesicle-mediated export and assembly of pore-forming oligomers of the enterobacterial clyA cytotoxin.";
Cell 115:25-35(2003).

MASS SPECTROMETRY, DISULFIDE BOND, AND MUTAGENESIS OF TVR-96; ASN-156;
TYR-164 AND ARG-260.
MEDLINE=20576278; PubMed=11006277; DOI=10.1074/jbc.M005420200;
Atkins A., Wyborn N.R., Wallace A.J., Stillman T.J., Black L.K.,
Fielding A.B., Hisakado M., Artymuk P.J., Green J.;
"Structure-function relationships of a novel bacterial toxin,
hemolysin E. The role of alpha G.";
J. Biol. Chem. 275:41150-41155(2000).

INDUCTION.
MEDLINE=20507803; PubMed=11053378;
Westermarck M., Oscarsson J., Mizunoe Y., Urbonaviciene J., Uhlin B.E.;
"Silencing and activation of clyA cytotoxin expression in Escherichia coli.";
J. Bacteriol. 182:6347-6357(2000).

INDUCTION.
MEDLINE=22053220; PubMed=12057949;
Spory A., Bosserhoff A., von Rhein C., Goebel W., Ludwig A.;
"Differential regulation of multiple proteins of Escherichia coli and Salmonella enterica serovar Typhimurium by the transcriptional regulator clyA.";
J. Bacteriol. 184:3549-3559(2002).

MUTANT PMW16 DEL.
MEDLINE=22829856; PubMed=12949101;
Wai S.N., Westermarck M., Oscarsson J., Jass J., Maier E., Benz R.,
Uhlin B.E.;
"Characterization of dominantly negative mutant clyA cytotoxin proteins in Escherichia coli.";
J. Bacteriol. 185:5491-5499(2003).

R-XRAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
MEDLINE=20123445; PubMed=10660049;
Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,
Green J., Artymuk P.J.;
"E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy.";
Cell 100:265-276(2000).

-I- FUNCTION: Toxin, which has some hemolytic activity towards mammalian cells. Acts by forming a pore-like structure upon contact with mammalian cells.
SUBUNIT: Monomer and oligomer. In periplasm, it is present as a monomer, while in outer membrane vesicles, it oligomerizes to form a pore structure that is active. Probably forms an octamer.
SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer membrane vesicles. Also found in the periplasmic space.

-I- INDUCTION: During anaerobic growth. Weakly or not expressed in most strains. It is activated by slyA, while it is silenced by HNS. Its expression is also regulated by CRP and FNR.
PTM: In periplasm, it forms a disulfide bond between Cys-86 and Cys-284, which prevents the oligomerization. In outer membrane vesicles, the redox status prevents formation of the disulfide bond, leading to oligomerization and pore formation.
MASS SPECTROMETRY: MW=34940; METHOD=Electrospray; RANGE=1-302; NOTE=Ref.12.

-I- SIMILARITY: Belongs to the hemolysin E family.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U57430; BAB07048.1; ALT_INIT.
EMBL; X98615; CAA67204.1; ALT_INIT.
EMBL; AJ001829; CAA05035.1; -.

DR EMBL; U73842; AAD04731.1; -.
DR EMBL; AS000216; AAC74266.1; ALT_INIT.
DR EMBL; D90751; BAA36016.1; ALT_INIT.
DR EMBL; D90752; BAA36029.1; ALT_INIT.
DR EMBL; D90753; BAA36037.1; ALT_INIT.
DR EMBL; AF240780; AAL55667.1; -.
DR EMBL; U22466; AAA92081.1; -.
DR EMBL; U13610; -; NOT_ANNOTATED_CDS.
DR PDB; 1QOY; X-ray; A=I-302.
DR EMBASE; EB3032; hlyE.
DR ECGene; EGI3243; hlyE.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
DR 3D-structure; Complete proteome; Cytolysis; Direct protein sequencing;
KW Hemolysis; Toxin; Transmembrane.
FT INIT_MET 0
FT TRANSHEM 182 202 Potential.
FT DISULFID 86 284 In monomeric form. CH9802).
FT VARIANT 174 174 K -> R (in strain CH9802).
FT MUTAGEN 200 200 G -> A (in strain CH9802).
FT MUTAGEN 87 89 GVA->DVD: Abolishes cytotoxic activity.
FT MUTAGEN 96 96 Y->H: Strongly reduces cytotoxic activity.
FT MUTAGEN 142 143 NA->DD: Abolishes cytotoxic activity.
FT MUTAGEN 156 156 N->H: Strongly reduces cytotoxic activity.
FT MUTAGEN 164 164 Y->C: Strongly reduces cytotoxic activity.
FT MUTAGEN 182 185 Missing: In PMW16; retained in cytosol.
FT MUTAGEN 182 183 AG->DD: Abolishes cytotoxic activity.
FT MUTAGEN 186 187 AG->DD: Abolishes cytotoxic activity.
FT MUTAGEN 260 260 R->K: Strongly reduces cytotoxic activity.
FT MUTAGEN 267 267 D->A: Strongly reduces cytotoxic activity.
Query Match 91.18; Score 1376; DB 1; Length 302;
Best Local Similarity 89.7%; Pred. No. 4.7e-83;
Matches 271; Conservative 18; Mismatches 13; Indels 0; Gaps 0;
OY 2 TGIFAEQTVVVKSAIETADGALDFYNYKLDQVDPKTFDETIKELSRFKQYSQASVL 61
DB 1 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVDPKTFDETIKELSRFKQYSQASVL 60
OY 62 VGDITKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKKSAAQKDLIRILD 121
DB 61 VGDITKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKKSAAQKDLIRILD 120
OY 122 DGVNKLNEAQSLLSGSSQSFNNASGKLLALDSQLTDFSEKSSYFQSDVRIRKEAYAGA 181
DB 121 DGVNKLNEAQSLLSVSSQSFNNASGKLLALDSQLTDFSEKSSYFQSDVRIRKEAYAGA 180
OY 182 AAGIVAGPFGIILISYIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDRAKL 241
DB 181 AAGIVAGPFGIILISYIAAGVIEGKLIPELNDRLKAVONFTSLSVTVKQANKDIDRAKL 240
OY 242 KLATETIAIGIKETETETTRFVVDYDDLMLSLGAAKMKMTNCNEYQORHGKKTLEVP 301
DB 241 KLTTETIAAIGETETETTRFVVDYDDLMLSLGAAKMKMTNCNEYQORHGKKTLEVP 300
OY 302 DI 303
DB 301 EV 302
RESULT 4
HLYE_ECO57
ID HLYE_ECO57 STANDARD; PRT; 302 AA.
AC Q9REB3
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).

GN Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=1944, ECS1677;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EHEC;
RX MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;
RA del Castillo F.J., Moreno F., del Castillo I.;
RT "Characterization of the genes encoding the SheA haemolysin in
Escherichia coli O157:H7 and Shigella flexneri 2a";
RL Res. Microbiol. 151:229-230(2000).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074915; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Toxin, which has some hemolytic activity towards
mammalian cells. Acts by forming a pore-like structure upon
contact with mammalian cells (By similarity).
CC -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a
monomer, while in outer membrane vesicles, it oligomerizes to form
a pore structure that is active. Probably forms an octamer (By
similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer
membrane vesicles. Also found in the periplasmic space (By
similarity).
CC -!- PTM: In periplasm, it forms a disulfide bond, which prevents the
oligomerization. In outer membrane vesicles, the redox status
prevents formation of the disulfide bond, leading to
oligomerization and pore formation (By similarity).
CC -!- SIMILARITY: Belongs to the hemolysin E family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AJ238954; CAB64962.1; ALT_INIT.
DR EMBL; AF005335; AAG56033.1; ALT_INIT.
DR EMBL; AF002555; BAB35100.1; ALT_INIT.
DR PIR; E85696; E85696.
DR PIR; E90838; E90838.
DR HSSP; P77335; 1QOY.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
KW Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
FT INIT MET 0 0 By similarity.
FT TRANSMEM 182 202 Potential.
FT DISULFID 86 284 In monomeric form (By similarity).
SQ SEQUENCE 302 AA; 33585 MW; F261E29E1DE5FC87 CRC64;

Query Match 90.1%; Score 1361; DB 1; Length 302;
Best Local Similarity 88.7%; Pred. No. 4.6e-82;
Matches 268; Conservative 20; Mismatches 14; Indels 0; Gaps 0;
QY 2 TGIFAEQTVVVKSAIETADGALDFYNYKLDQVFPWKTFDETIKELSRFKQYSQASVYL 61
DB 1 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVFPWKTFDETIKELSRFKQYSQASVYL 60
QY 62 VGDIKVLMDSDQKYFEATQTVYEWCVVTVOLLISAYILLFDEYNEKKASAKDILIRILD 121
DB 61 VGNIKTLMDSDQKYFEATQTVYEWCVVTVOLLISAYILLFDEYNEKKASAKDILIRILD 120
QY 122 DGVNKLNEAQSLLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKREYAGA 181
DB 121 DGITKLNKAQSKLLVSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKREYAGA 180
QY 182 AAGIVAGPFGILLISYISIAAGVIEGKLIPELNDRLKAVQNFETSLVTVKQANKDIDAAL 241
DB 181 AAGVAGPFGILLISYISIAAGVVEGKLIPELNKLVKLSVQSFETSLVTVKQANKDIDAAL 240
QY 242 KLAIEIAAIGIKETETETTRFVVDYDDLMLSLKGAKKMINTCNEYQORHGKKTLLFVP 301
DB 241 KLTIEIAAIGIKETETETTRFVVDYDDLMLSLKGAKKMINTCNEYQORHGKKTLLFVP 300
QY 302 DI 303
DB 301 EV 302
RESULT 5
QX2S8 PRELIMINARY; PRT; 300 AA.
ID Q9X2S8
AC Q9X2S8;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-2004 (TRENBLrel. 26, Last annotation update)
DE Hemolysin.
GN Name=hlyE;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=99242013; PubMed=10227474;
RX Reingold J., Starr N., Maurer J., Lee M.D.;
RA "Identification of a new Escherichia coli She haemolysin homolog in
avian E. coli";
RL Vet. Microbiol. 66:125-134(1999).
DR EMBL; AF052225; AAD28079.1; -.
DR HSSP; P77335; 1QOY.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
SQ SEQUENCE 300 AA; 33555 MW; 6580B66C44A7B4BC CRC64;
Query Match 73.1%; Score 1104; DB 2; Length 300;
Best Local Similarity 73.4%; Pred. No. 4.5e-65;
Matches 213; Conservative 39; Mismatches 38; Indels 0; Gaps 0;
QY 6 AEQTVVVKSAIETADGALDFYNYKLDQVFPWKTFDETIKELSRFKQYSQASVYLVDGI 65
DB 4 ADQIVETVKTATDADKALDLYNKYLDQVFPWKTFDETIKELSRFKQYSQASVYLVDGI 63
QY 66 KVLMDSDQKYFEATQTVYEWCVVTVOLLISAYILLFDEYNEKKASAKDILIRILDGWN 125
DB 64 KSLMLNSQDRYFEATQTVYEWCVVTVOLLISAYILLFDEYNEKKASAKDILIRILDGWN 123
QY 126 KLNBAQSKLLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKREYAGAAAGI 185
DB 124 KLEKAAQSLHASSQSFNSASGKLLALDSQLTNDPSEKSSYFQSQVDRIKREYAGAAAGV 183

CC modified and this statement is not removed. Usage by and for commercial

```

QY 211 LNDRLKAVQNFSTLSVTVKQAKDIDAALKLATEIAAIGIKTETETTRFYVDYDDL 270
DB 1 MNKLKLSALSPFTLSNTVKQAKDIDAALKLATEIAAIGIKTETETTRFYVDYDDL 60

QY 271 LSLKGAAGKAMTNCNVOORHGKTKLLVPDI 303
DB 61 LSLKGAAGKAMTNCNVOORHGKTKLLVPDI 93

RESULT 8
Q9LIW7 PRELIMINARY; PRT; 895 AA.
AC Q9LIW7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Similar to an Arabidopsis thaliana chromosome BAC genomic
DE sequence.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001111; BAA90502.1; -.
DR Gramineae; Q9LIW7; -.
SQ SEQUENCE 895 AA; 100520 MW; BA4F7EFD70FFA3 CRC64;

Query Match 9.0%; Score 136.5; DB 2; Length 895;
Best Local Similarity 22.4%; Pred. No. 1.6;
Matches 77; Conservative 67; Mismatches 121; Indels 79; Gaps 17;

QY 8 QTVVVVKAIAETADGALD-----FYNKYLQDVIPWTKTFTDIK-----ELSRFKQRY 54
DB 274 ERVEITLSSEVRLKGLDSTAESSESKNRETELV--KNLESEVSLVKGLFEARIIER 331

QY 55 SOEASVLGDIKVLWDSQDKYFEATQTVYEW---CGVVQTLLSAVILLFQY-NEKAS 110
DB 332 LAETEKLEELKSEVADAKAASEARQIFEEWKHAGLLEMELEA-VTLDKFKGSELAS 390

QY 111 AQDKILIRILDGVNKLNEAQSLLGSSOSFNASGKLLALD---SQTNDPFSKSYF- 166
DB 391 TTEE-----LQKIQSALQDRSEIEVLKGTALIEVARLLADVNESNEQFD 438

QY 167 -----QSQVDRIRKEAYAGAAGIVAGPFGGLIISYSIAAGV-IEG-----KLPIE 210
DB 439 ASQOEVFGLQTTIDVLRNKLAEAEBASEA-----LNNEKAANYKIEGLTEENVKLISE 492

QY 211 LND-----RLKAVQNFSTLSVTVKQAKDIDAALKLATEIAAIGIKTET 257
DB 493 LNETDRBEKEKRAVEDLTAALUS-----ESSDKAKEAHERVLSKEDDHEHALAIGOLKMAL 549

QY 258 ETTR----FYVDYDDLMLSLKGAAGKAMTNCNVEYQOR-HGKKT 296
DB 550 KSTKSEYEWLDEANYDITCLARKNVKLEAEVKNKYRESCEKSET 593

RESULT 9
P71497 PRELIMINARY; PRT; 495 AA.
AC P71497;
DT 01-FEB-1997 (TRENBLrel. 02, Created)
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE ClpB (Fragment).
GN Name=clpB;
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2095;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25416;
RX MEDLINE=97148974; PubMed=8995799;
RA Falah M., Gupta R.S.;
RT "Phylogenetic analysis of mycoplasmas based on Hsp70 sequences:
RT cloning of the dnaK (hsp70) gene region of Mycoplasma capricolum.";
RL Int. J. Syst. Bacteriol. 47:38-45(1997).
DR EMBL; U51235; AAB09427.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001270; Chaperin_c1pA/B.
DR PRINTS; PRO0300; CLPPTROTEASEA.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding.
FT NON TER 1
SQ SEQUENCE 495 AA; 56358 MW; A194DD51FACDF8D0 CRC64;

Query Match 8.7%; Score 132; DB 2; Length 495;
Best Local Similarity 24.3%; Pred. No. 1.5;
Matches 79; Conservative 47; Mismatches 115; Indels 84; Gaps 16;

QY 2 TGIFAEQTVVVVKAIAETADGALDGFYNKYLQDVIP-----WKTDETIKELSR 49
DB 178 TGIWVVKLISSEKERLLNLEDLLKKYVKGQDAIKAVTSAIMRSRGIKNPKPIGSLF 237

QY 50 F-----KQEYSQE-ASVLVGDIKVLWDSQDKYFEATQTV-----YEWCGVVT 91
DB 238 FGPTGVGKTEVARSLADILFNSPKMHRLDSEYMEKHSIAKLICAPPYGYEGGRLT 297

QY 92 QLL--SAY-ILLFDEYNEKKAQAQKILIRILDGVNKLNEAQSLLGSSOSFNASGKL 148
DB 298 EAVRNPYSIILFDEI-EKAHSDVFNILLOILDDG--RLTDS-----LCKTIDFKNT---I 347

QY 149 LALDSQLTNDPFSKSYF-----QSQVDRIRKEAYAGAAGIVAGPFGGLIISYSIAA 200
DB 348 IVMTSNIAQYLLTSDDDFVQDDQKIQAEQLNTPFPEFLNRIDNIVY--FNALSQVTI-- 403

QY 201 GVIEGKLIPENDRLKAVQNFSTLSVTVKQAKDIDAALKLATEIAAIGIKTETET 260
DB 404 GEIVDKLDELITRLQDQEQNYFINFS-----EARNKIINE----- 439

QY 261 RFYVDYDDL--LSLLKGAAGKAMT 284
DB 440 ---GYDLFGARLLKDLIKLIET 460

RESULT 10
Q7ROB6 PRELIMINARY; PRT; 2723 AA.
AC Q7ROB6;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE 235 kDa rhostry protein.
GN Name=py01185;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteza M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,

```

RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
parasite Plasmodium yoelii yoelii";
RL Nature 419:512-519(2002). shown here is derived from an
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL: AABL01000312; EAA20475.1; --
DR InterPro: IPR011561; Pox I1 rel.
DR InterPro: IPR006499; ReticulocyteBP.
DR ProDom: PD014111; Pox_I1 rel.; 1.
DR TIGRFAMs: TIGR01612; 235kDa-fam; 1.
SQ SEQUENCE 2723 AA; 320451 MW; 08CD9C2EF7672ED2 CRC64;

Query Match 8.6%; Score 130; DB 2; Length 2723;
Best Local Similarity 19.7%; Pred. No. 16;
Matches 62; Conservative 54; Mismatches 96; Indels 102; Gaps 10;

QY 11 EVVKSATETADGALDFYNYKLDQVLPWKTFDTIKELSRFKQE-----53
Db 1464 EELKGYDKSGYKDEVNKNT-----KTIENKKELEKYKNEAINLNKFSFMLKNKI 1517

QY 54 --YSQASVVGDIKVLMDSDQKYFEATQTVYVCGVVTQLLSAYILLFDEYNEKKASA 111
Db 1518 IOTKKDLLEIINEIKWHEK-----ILQADKSEQKINT 1551

QY 112 QKDILIRLDDGVNKLNEAKSLGSSQSFNNASGKLALDSQLTNDPSEKSSYFQSQVD 171
Db 1552 IKNEQIQKDDTTSN-DKSNKAIIGIQASLDKFKFLKI-----NDIRTKSNFLKETE 1605

QY 172 RIRKEAYAGAAGIVAGPGLIISYSIAAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQ 231
Db 1606 NIEKQ-----ISNLSINSQEIK-----LKNEIDLNTLQKFESL-----KD 1642

QY 232 ANKQIDAACKLKLAT-----ETAAIGEIKETETETTFYVYDDLMLS 272
Db 1643 QKQVIEDQKTELNDPSEIENIESDVSNQKNVEMGILIEKIKENADTNKNQIESTK---E 1699

QY 273 LKGAAGKMNNTON 286
Db 1700 LKPTIENLISFN 1713

RESULT 11
Q6MT03 PRELIMINARY; PRT; 713 AA.
AC Q6MT03;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE ATP dependant protease ClpB.
GN Name:clpB; OrderedLocusNames=MSC_0613;
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG1;
RX PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundeberg J.,
RA Johansson K.-E., Pettersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
strain PG1T, the causative agent of contagious bovine pleuropneumonia
(CBPP).";
RT Genome Res. 14:221-227(2004).
RL EMBL: BX842644; CAE77235.1; --
DR GO:0008233; F:peptidase activity; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase.Centr.
DR InterPro: IPR001270; Chaprinin_c1pA/B.
DR Pfam: PF00004; AAA; 2
DR PRINTS: PR00300; CLPPTOTASEA.

DR SMART: SM00382; AAA; 2.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding; Complete proteome; Protease.
SQ SEQUENCE 713 AA; 80782 MW; 1B5D204A9E29BE50 CRC64;

Query Match 8.5%; Score 128.5; DB 2; Length 713;
Best Local Similarity 24.5%; Pred. No. 4;
Matches 81; Conservative 48; Mismatches 124; Indels 77; Gaps 17;

QY 2 TGIFAEQTVVVKSAIETADGALDFYNYKLDQVLPWKTFDETI-----KELSRF 50
Db 396 TGIIVDRLLISSEKEKLNLEDLLKKYVKGQDQAI--KAVTSAIMRSRGINKPDKPIGSF 453

QY 51 -----KQYISOE-ASVLVGDIKVLMDSDQKYFEATQTV-----YEWCGV 89
Db 454 LFLGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMEKHSVAKLIGAPPVGVYEEGGR 513

QY 90 VTQLL--SAY-ILLFDEYNEKKASAQKDIILIRLDDGVNKLNEAKSLGSSQSFNN---143
Db 514 LTAVERNRPYSIVLFDEI-EKAHTDVFNILLQILDDG--RLTDS-----LGTIDFKNTII 566

QY 144 -----ASGKLALDSQLTNDPSEKSSYFQSQVDRIKKEAYAGAAGIVAGPGLIISYS 197
Db 567 VMTSNIASQVLLSDELQVD-----DQIQEELNKVFRPEFLNIDNIVY--FNALSQVT 620

QY 198 IAAGVIEGKLIPELNDRLKAVQNFSTLSVTVKQANKDIAAKLKL--ATEIAAIGEIKT 255
Db 621 I--GEIVDKVLEELSTRLQDEQNVFINFSEAR--NKIINEGYDRLFGARPIKRYIEKNI 676

QY 256 ET-----ETTRFYVDYDDLMLSL 273
Db 677 ETLIAHYIISGEVVVENTRYLIDVKNNOFTL 706

RESULT 12
CAE77235 PRELIMINARY; PRT; 713 AA.
AC CAE77235;
DT 02-MAR-2004 (Tremblrel. 27, Created)
DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)
DT 13-APR-2004 (Tremblrel. 27, Last annotation update)
DE ATP dependant protease ClpB.
GN CLPB OR MSC_0613.
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG1;
RX PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundeberg J.,
RA Johansson K.-E., Pettersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
strain PG1T, the causative agent of contagious bovine pleuropneumonia
(CBPP).";
RT Genome Res. 14:221-227(2004).
RL EMBL: BX842644; CAE77235.1; --
DR Protease.
KW SEQUENCE 713 AA; 80782 MW; 1B5D204A9E29BE50 CRC64;

Query Match 8.5%; Score 128.5; DB 2; Length 713;
Best Local Similarity 24.5%; Pred. No. 4;
Matches 81; Conservative 48; Mismatches 124; Indels 77; Gaps 17;

QY 2 TGIFAEQTVVVKSAIETADGALDFYNYKLDQVLPWKTFDETI-----KELSRF 50
Db 396 TGIIVDRLLISSEKEKLNLEDLLKKYVKGQDQAI--KAVTSAIMRSRGINKPDKPIGSF 453

QY 51 -----KQYISOE-ASVLVGDIKVLMDSDQKYFEATQTV-----YEWCGV 89
Db 454 LFLGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMEKHSVAKLIGAPPVGVYEEGGR 513

QY 90 VTQLL--SAY-ILLFDEYNEKKASAQKDIILIRLDDGVNKLNEAKSLGSSQSFNN---143

```

Db 514 LTEAVRRNPYSVLDFEI--EKARTDVFNILLQLDDG--RLTDS-----LGKTFIDFKNTII 566
Qy 144 -----ASGKLALDSQLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGGLIISVS 197
Db 567 VMTSNIASQYLTSBELVQVD-----DQKIQEELNKVFRFELNRIDNIVY--FNALSQVT 620
Qy 198 IAAGVIEGKLIPELNDRKAVNQFTSLSVTVKQANKOIDAAKJL--ATEIAAIGEIKT 255
Db 621 I--GBIVDKVLELSTRLQDEQNYFINFSEEAR--NKIINEGYDRLFGARPIKRYIEKNI 676
Qy 256 ET-----ETTFYVDYDDLMLSL 273
Db 677 ETLIAHYIISGEVVENTRYLIDVKNQFTL 706

RESULT 13
Q7N5U5 PRELIMINARY; PRT; 577 AA.
AC Q7N5U5;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Similar to putative phage protein.
GN OrderedLocNames=plu1837;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusnok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-P.,
RA Dassa E., Deroose R., Derzelle S., Freyssinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; BX571865; CAE14130.1; -.
DR PhotocList; plu1837; -.
KW Complete proteome.
SQ SEQUENCE 577 AA; 65287 MW; B7F86BE4B34BEDE1 CRC64;

Query Match 8.4%; Score 126.5; DB 2; Length 577;
Best Local Similarity 19.0%; Pred. No. 4.2; Mismatches 155; Gaps 14;
Matches 80; Conservative 55; Indels 132; Indels 155; Gaps 14;

Qy 1 MTGIPAEQTVVVK-----SAIETADGALDFYNKLDQV-----IPWKT----- 40
Db 103 ITGDFSKGFSDLVNFEMKLPNREGDNLTPPPAYVFLPFIQIKSWSPWNSFENLQY 162
Qy 41 -----DETIKLSRFKQYSEASVLVDIKV----- 68
Db 163 SNWKGPIKYFTGYLKPPEHFOEENIYEYSEIKESAHKIEKFSQAVEIVDNTVDSPT 222
Qy 69 -----LMSDQKYPEATQTVYEMGVTVTLLSAYILLFDEYNEKKA 109
Db 223 LQNDFFIKQNEIKNELYDLIDYQRLYDAQTI--TSNIYDLESQYALAITSALEEA 279
Qy 110 -----SAQKDL-----TRILDDGVNKLNEAQKSLGSSQSPNNASG 146
Db 280 DYKFAVESIPTDYELCPICGLTHDLSLNPRLALLSEKDSLINEA-----NSTAS 328
Qy 147 KLLALDSQLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGGLIISYIAAGVIEGK 206
Db 329 KIVELKSSL-NSLNEDAQFITNEIERINNK-----YITDDAVKKG 368
Qy 207 LIPELNDRL-----KAVNQFTSLSVTVKQANKDI-----DAAKL----- 241
Db 369 LIAQVIDTLSTENVSKNIQIKIDNEDLNIISKANSISKELKKDQKLLSTKEELNSFM 428
```

```

Qy 242 -KLATEIAAIGEIKTETETTRFYVDYDDLMLSLLKGAARKMI--NTCNEYQQRHGKKTLL 299
Db 429 SKLLANIEALSGTGINLSKVSPDQYKQLLGGAAGAAAGLLAYQLSVLQIHGAKTCVV 488
Qy 300 VP 301
Db 489 PP 490

RESULT 14
QSG015 PRELIMINARY; PRT; 1521 AA.
AC QSG015;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=BQ05330;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RX PubMed=15210978;
RA Alsmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
RA Canbaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,
RA La Scola B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721 (2004).
DR EMBL; BX897700; CAP26028.1; -.
DR InterPro; IPR009074; Apolipo_A_E_C3.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR010989; t-snare.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1521 AA; 169742 MW; 12CC93BCB2C97920 CRC64;

Query Match 8.3%; Score 125; DB 2; Length 1521;
Best Local Similarity 18.4%; Pred. No. 17; Mismatches 147; Indels 90; Gaps 8;
Matches 69; Conservative 69; Indels 147; Indels 90; Gaps 8;

Qy 4 IFAEQTVVVKSAIETADGALDFYNK-----YLDQVIPWK-----TFDET 43
Db 928 VLSEQTACTVESFTASHNAQTLNETHTSATAIEEVLNERCNVLHSMQNLKNLGVQ 987
Qy 44 IKELSRFKQYSEASVLVDIKVLLMDSQDKYFEATQTVYEMGVTVTLLSAYILLFDE 103
Db 988 LSDVSNHLEEKQTATQISGHVEKLTETLVNLAQAQNTTESISHLTQHSIQSLSLSTQ 1047
Qy 104 YNEKASAKQDILIRIL-----DDGVNKLNEAQKSLGSSQSPNNASGKL 148
Db 1048 DAQRIVYAQNESLNSLTQTNSLTQVTAMKEDLVNNISILKQLNQSISYFHSNML 1107
Qy 149 LALDSQLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGGLIISYIAAGVIE-- 204
Db 1108 LSTVQNDIQSEFANNFRFTNQ-----AAEHLASNAQNLNNVTVLQGLSQNIFE 1159
Qy 205 -----GKLIPELNDRLKAVNQFTSLSVTVKQANKOIDAAKJLATEIAAIGEI-- 253
Db 1160 KIGHTSTFGSHAKTLSETIHILEKSENSLTLEEKHKLTSALSSALVSKSNEINKLIE 1219
Qy 254 -----KTET-----TFYVDYDDLMLSL--LKAARKM 281
Db 1220 YYENVLSLAFERTDTNTRNSTHSLQQSLNQLINEASTRFGAAEDIRKSADEIRSELSKI 1279
Qy 282 INTCNEYQQRHGKKT 296
Db 1280 NNDINESVQNLPEKT 1294
```



```
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01001582; EAA17018.1; -.
DR InterPro: IPR006499; ReticulocyteP.
DR TIGRFAMs: TIGR01612; 235kDa-fam; 1.
FT NON TER 2664
SQ SEQUENCE 2664 AA; 312670 MW; FDE72DB05743F4AB CRC64;

Query Match 8.0%; Score 121; DB 2; Length 2664;
Best Local Similarity 21.3%; Pred. No. 60;
Matches 71; Conservative 67; Mismatches 140; Indels 56; Gaps 14;

QY 8 QTVEVVKSAIET--ADGALDFYNKYLDQVTPWK-TFDETIKELSRPK---QEYSQASVL 61
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1457 QYIIATKEDLTHNDHSNISKNEHKDKSKAYKQDEIDNNIKKLEENKKIPEKYQETTVL 1516

QY 62 VGDIKVLMDSQDKFEATQTVYEWCGVVTQLLSAY--ILLDFEYNEKKAQAQKILIRI 119
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1517 LNKYSALAKNK--FDQTKDSEL--IIKEIKDHKNCVVQSETSEQKMTIKNEQIFI 1571

QY 120 LDDGVN--KLNFAQKSLGSSQSNFASGKLLADS--QLTNDPSEKSSYFQSVQDIRK 175
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1572 DSNNTNNDKSNNAIISIQSIEKFKT---KLLNNGIMQKNDCLKETTIDIEQISKU-- 1626

QY 176 EAYAGAAAGIVAGPFLGIISYSIAAGVIEGKL-----IPELNDRLKAVQNFTSLV 227
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1627 -SINNQDTFLVNNETKLNDFKFLSLAKEKKIYIQDREKELNEVNSQIKNIEN----- 1678

QY 228 TVQANKDIDAKLKLATEIAAIGIKETETTRTFYVDYDMLSL-----L 274
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1679 DVNQYKDYEGIVEKINEIAKANKDKIESTKESTIEPIIIONIMSFNSDIEDIASNEAL 1738

QY 275 KGAAKMINTNEVQQRHG-----KKTLLVDPDI 303
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1739 ETYKTSMNNTNEFMESYDLIKKYSETVSKPTI 1772

RESULT 18
ID Q7M918 PRELIMINARY; PRT; 693 AA.
AC Q7M918;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN OrderedLocusNames=WS0886;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ 1740;

RX MEDLINE=22882897; PubMed=14500908;
RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
DR EMBL: BX571659; CAE09993.1; -.
KW Complete proteome.
SQ SEQUENCE 693 AA; 77641 MW; CB57A1F23CFACCC2 CRC64;

Query Match 8.0%; Score 120.5; DB 2; Length 693;
Best Local Similarity 21.3%; Pred. No. 13;
Matches 64; Conservative 52; Mismatches 124; Indels 61; Gaps 9;

QY 11 EVVKSAIETADGALDFYNKYLDQVTPWKTFDETIKELSRPKQEYSQASVLVGDIKVLM 70
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 EEVTSSTKEAQESLTQSSRWLQERIQ-----SDISLAQQHFTFHLGGLSTEFKALV 367

QY 71 DSQKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKAQAQKILIRILDDGVNKLNEA 130
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 SLREQLEGANEVLSLLQKSSQESAFLL----EQEKMAKESTLVLFKAKKERFERLSEA 423

QY 131 -----OKSLGSSQSENN-----ASGKLLALDSLTNDPSEKSSYFQSV 170
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 FLTONKGALEGNALDFRGFWEEYATRWRESSELQGTILLETNQVRASFAELSEGVLQN 483

QY 171 DRIRKEAYAGAAAGIVAGPFLGIISYSIAAGVIEG--KLIPELNDRLKAVQNFTSLV 227
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 SAMGEQIKAG-FEGMTQALSGMLAKALEGGIDGVKSVSENNQTL----- 530

QY 228 TVQANKDIDAKLKLATEIAAIGIKETETTRTFYVDYDMLSLGKAQKAKMINTCNE 287
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 ----AKETLQTNHSQSLGLLGESTTAIQTR-----LLEEGAGKSLKLNLALE 577

QY 288 Y 288
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 Y 578

RESULT 19
ID Q7VI37 PRELIMINARY; PRT; 555 AA.
AC Q7VI37;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=HH0771;
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Froesch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL: AE017146; AAP77368.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 555 AA; 61077 MW; 1E2515BCB83A804 CRC64;

Query Match 7.8%; Score 117.5; DB 2; Length 555;
Best Local Similarity 22.9%; Pred. No. 16;
Matches 83; Conservative 54; Mismatches 122; Indels 103; Gaps 18;

QY 13 VKSAIETADGALDFYNKYLDQV-----IPWKTFDETIKELSRPKQEYSQASVLV 62
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 13 VKSAIETADGALDFYNKYLDQV-----IPWKTFDETIKELSRPKQEYSQASVLV 62
```

```
Db 6 VKFTFNKVGSLQKIKKELDSINKNSANELYALBLKHEQKMEHI-----QEYSKAYKDLQ 61
QY 63 GDIVKLLMDS-----QKYPBATQVYEWGVVTPOLLISAVI--LLFDEYN 105
Db 62 SDIQSALNEDQIGFNGVHYFRSLINELFSLQSVIT-KGFATSIAGAFWESSVIESMN 120
QY 106 EKASQAQDIILIRLDDGVNKLNEAQK-----SLLGSSQS---FNNASGKLLA--LDSQ 154
Db 121 KSLASAIQSLFSLNFD-KNRIANVQKEDWHLSEVVGAAFAFGIGVAGGLVGNFLGDE 179
QY 155 LTNDSEKSSYFQSQVDRIRKEAVAGAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDR 214
Db 180 VNAQKTQKSA-----NNGAMAGAAAGRAIGSFVPIGTSL--GGLVGLVGLGGT 228
QY 215 LKAVQNFSTLSVT-----VKQANKOIDAAKLKLATEIAAIGK-----TETETTR 261
Db 229 LFGSFN-STKLTTTAQGVLLISKATKDNVSAR-----EPADKEEVKKKMWGLQSNSTTWK 282
QY 262 FYVD-----YDDL-----LSLLGAAKMMINTCNEYQQRHGKKT 297
Db 283 EYDASNFAKGIQSIIRGYEYLLQDIGSVKLSIAKGNYSYADILNA-----GAKEL 337
QY 298 LE 299
Db 338 IK 339

RESULT 20
Q07569 PRELIMINARY; PRT; 2139 AA.
ID Q07569; O02504;
AC Q07569; O02504;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin heavy chain.
GN Name=hmhCA;
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=93295430; PubMed=8515774;
RA Raymond-Denise A., Sansonetti P., Guillen N.;
RT "Identification and characterization of a myosin heavy chain gene
RL (hmhCA) from the human parasitic pathogen Entamoeba histolytica.";
RN Mol. Biochem. Parasitol. 59:123-131 (1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RA Guillen N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L03534; AAB48065.1; -.
DR PIR; T18296; T18296.
DR HSSP; P08799; 1MND.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;

Query Match 7.7%; Score 117; DB 2; Length 2139;
Best Local Similarity 19.0%; Pred. No. 85;
Matches 75; Conservative 47; Mismatches 114; Indels 158; Gaps 13;

QY 7 EQTVVVKSAIETADGALDFYKLDQVTPKTFDETIKELSRPKQYSQASVYLVGDIK 66
Db 965 EVELTNSQINTLNATVD-----KDKTIAEQESIDEKEDEITKLGDIK 1011
```

```
QY 67 VL-----LMSODKVFETATQ 81
Db 1012 LLEEKDDLEQRADVSATKDIKAKLNKLTIECEDAKETIAKLEQLEDEENKKNKDLTN 1071
QY 82 TVYEW---CGVTVQLLSAYILLFDEYNEKASAKQDIILIRLDDGVNKLN-----E 129
Db 1072 ELQOTQLKGETEKSLAAQVAA-----TKASDERDTLSQNLN--EKLTTKNLTKTKAD 1124
QY 130 AKSLIGSSQS-----NNASGKLLALDSOLTNDSEKSSYFQSQVDRIRK 175
Db 1125 LEKISGLKQDYEDLEDDKKNKIEGDLRNAORKIKELDEITKG-ADVSQYLQK-----K 1178
QY 176 EAVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNRLKAVQNFSTLSVTVQANKD 235
Db 1179 EYESQ-----IAKMQEKEAIGNDVKNKERTIKEKELE 1212
QY 236 I-----DAAKLK--LATEIAAIGEIKTETETTRFYVDYDML----- 271
Db 1213 IQSLQEKLDTEVEKEDAEKKKEIKEMKALQEEKENVENSSNSTEKKKKLEDNKDT 1272
QY 272 -----SLLGAAKMMINTCNEYQQRHGK 294
Db 1273 QKLLDWTADNEKLIKAKAKDLEAQLNEVDNHEK 1306

RESULT 21
Q064067 PRELIMINARY; PRT; 478 AA.
ID Q064067;
AC Q064067;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein yonD.
GN Name=yonD;
OS Bacteriophage SPBC2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RC Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.;
RA Karamata D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC13027.1; -.
DR PIR; T12818; T12818.
KW Hypothetical protein.
SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 478;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 58; Conservative 39; Mismatches 74; Indels 81; Gaps 12;

QY 25 DFYKRY-LDQVTPKTFDETIKELSRPKQYSQASVYLVGDIKVLMSODKVFETATQV 83
Db 268 DVIDTYFIVNVVWS--DENSVD-KYFKFNVTR-----TGDVTSIDFDSKTEVF---MT 315
QY 84 YEWCGVTVQLLSAYILLFDEYNEKASAKQDIILIRLDDGVNKLNEAQKSLIGSSQSFPNN 143
Db 316 RNWEEVPEPIQSOL-----NQKDEIQKLTQVQNQIN---KDKVGLIEQEQNT 360
QY 144 ASGKLLALDSQ-----LTNDFSEKSSYFQSQVDRIRKEAYAGAAAGIVAG 188
Db 361 ASEKLVQLNSEVVEQLKPYKKEHKTLLQKLSKNEFYKAKFEALNAEE----- 409
QY 189 PFGLIISYSIAAGVIEGKLIPELNRLKAVQNFSTLSVTVQANKOIDAAKLKLATEIA 248
Db 410 -----KFS-----TBEVQNL-----IHASVKQ--DEEGEKAVLQNTMLV 442
QY 249 AIGEIKTETETT 260
Db 443 DIVSVPTETNTT 454
```



```

RESULT 22
ID O31954
AC O31954;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE YonD protein.
GN Name=yonD; OrderedLocusNames=BSU21130;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=980404033; PubMed=9384377; DOI=10.1038/36786;
RA Kunet F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerth I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft I., Ehrlich S.D., Emmerson P.T.,
RA Enlian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerz-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seter S.J., Serror P.,
RA Shin B.S., Soldo B., Sorokin A., Taccioni E., Takagi T., Takahashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vasarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
DR EMBL; Z99115; CAB14031.1; -.
KW Complete proteome.
SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 478;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 58; Conservative 39; Mismatches 74; Indels 81; Gaps 12;

Qy 25 DFYKY-LDQVTPKTFDETIKELSRFKQVYSQASVLGDIKVLMDSDQKYFEATQV 83
Db 268 DVIDYFIVNYSWS--DENSDV-KYFKFNTR-----TGDVTSIDFSKTEVF---MT 315

Qy 84 YEWCGVTTQLSAYILLFDEYNEKKAQAQDILIRLDGQVKNLNEAQSLGSSQSPNN 143
Db 316 RNWEEVPEPIQSGL-----NQDQIKDLTKQVQNIN---KDKVGIEQFNT 360

Qy 144 ASGKLLALDSQ-----LINDSEKSSYFQSDVRIRKEAYAGAAAGIVAG 188
Db 361 ASEKVLQNLNSEVQELKPYKEKHEKTLLEQKSEKNEFYKAKFEALNAEE----- 409

Qy 189 PFGLIISYIAAGVTEGKLIPELNDRLKAVQNFFTSLSVTKQANKDIDAAKLKLAIEIA 248
Db 410 -----KFS-----TEEVQNL-----IHASVKQ-DEGEKAVQLQNTLVL 442

Qy 249 AIGEIKTETETT 260

```

```

Db 443 DLVSVPTETNTT 454

RESULT 23
Q7CMF0 PRELIMINARY; PRT; 652 AA.
ID Q7CMF0
AC Q7CMF0;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE S-layer protein, (PX01-90).
GN Name=EXA0124;
OS Bacillus anthracis str. A2012.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=191218;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2012;
RX MEDLINE=22061436; PubMed=12004073;
RA Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,
RA Holtzapple E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,
RA Keim P., Fraser C.M.;
RT "Comparative Genome Sequencing for Discovery of Novel Polymorphisms in
RT Bacillus anthracis";
RL Science 296:2028-2033(2002).
DR EMBL; AE011190; AAM36077.1; -.
DR InterPro; IPR001119; SLH.
DR InterPro; IPR010989; t-snare.
DR InterPro; IPR010978; tRNA_binding_arm.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH_DOMAIN; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

Query Match 7.7%; Score 116.5; DB 2; Length 652;
Best Local Similarity 19.0%; Pred. No. 22;
Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

Qy 25 DFYKYLDDVTPKTFDETIKELSRFKQVYSQASVL-----VGDIKVLMDSDQKYFE 78
Db 237 DVINQKIDE---FDKLSQRKDLERMLLELNQKLSQLKQSPQLDNLKNEKESQRLLE 293

Qy 79 ATQ-----TVYEWCGVTTQLSAYILLFDEYNE--EKASAQKDILIRLDGQVKN 126
Db 294 LNKXDSNRLELNSEIKKLNDRKAEILLSLIMELIKQSEFDKKIKNEKDDLKRELLNR 353

Qy 127 LNEAQK-----SLGSSQSPNNASGKLLALDSQLTNDSEKSSYFQ---SQV 170
Db 354 IAESKELAKKKAELNLTKLVELFKVQEAALNKKSQVLYYINKLDNELRELADKYKNSDKI 413

Qy 171 DRIRKEAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIPELNDRLKAVQNFFTSLSVTK 230
Db 414 SRLKNH-----IGYKNQKLEKLEN-----ELK 435

Qy 231 QANKDIDAAKLKLAT-EIAAIGEIKTETETTRFFVYVDL-----MLSLKGAAKMINT 284
Db 436 ECNKIDNTKKQALAEFDKSNKKQQLSESELVQLNKKIDELGKRHKHQRQLEASQKALDE 495

Qy 285 CNEYQQRHGKK 295
Db 496 AKENKKLAEK 506

RESULT 24
Q9X360 PRELIMINARY; PRT; 652 AA.
ID Q9X360
AC Q9X360;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE PX01-90 (S-layer protein,).

```

GN OrderedLocusNames=GBAA_px01_0124;
 OS Bacillus anthracis.
 OC Plasmid virulence plasmid px01, and plasmid px01.
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1392;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Strom N.A.
 RX MEDLINE=99445483; PubMed=10515943;
 RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
 RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,
 RA Martinez Y., Ricke D., Svensson R., Jackson P.J.;
 RT "Sequence and organization of px01, the large Bacillus anthracis
 RT plasmid harboring the anthrax toxin genes";
 RL J. Bacteriol. 181:6509-6515(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ames / isolate 0581; PLASMID=px01;
 RX MEDLINE=99445483; PubMed=10515943;
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
 RA Fraser C.M.;
 RT "Bacillus anthracis comparative genomics";
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF065404; AAD32394.1; -;
 DR EMBL; AE017336; AAT28865.2; -;
 DR PIR; B59102; B59102.
 DR TIGR; GBAA_px01_0124; -;
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF00395; SLH; 3.
 DR PROSITE; PS01072; SLH DOMAIN; UNKNOWN_1.
 KW Complete proteome; Plasmid.
 SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

 Query Match 7.7%; Score 116.5; DB 2; Length 652;
 Best Local Similarity 19.0%; Pred. No. 22;
 Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

 QY 25 DFYKYLQVWPWTFTETIKELSRFKQEQYSQASVL-----VGDIKVLMDSQDYFE 78
 DB 237 DVINQKIDE---FDKLSQRKDLERMLELNQKLSQLKQSPQLQDLNKLKESQSRLE 293

 QY 79 ATQ-----TVYEWGCVVTVLLSAYILLFDYNE--KASQAQKDLIRILDGYNK 126
 DB 294 LNKXDSNRLELNSEIKLNDRAKELLSLIMELIKQSEFDKIKNEKDDLNKREDLINR 353

 QY 127 LNEAQK-----SLGSSQSFNNASGKLALDSQLTNDSEKSSYFQ---SQV 170
 DB 354 IAESKELAKKKAELNTKLVELFKVQEAALNKKSGQYLYINKLDNELRELADKYKNSDNKI 413

 QY 171 DRIRKEAVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVK 230
 DB 414 SRLKNH-----IGEYNKQLEKIE-----ELE 435

 QY 231 QANKOIDAAKLKAT-ETAAIGEIKTETTFYVDYDDL-----MLSLKGAAKMINT 284
 DB 436 ECNKKIDNTTKQLAEFDKSNKKQOESELSVLQNKIDELGKRHKHQELESQSKALDE 495

 QY 285 CNEYQQRHGKK 295
 DB 496 AKEINKKLAEK 506

 RESULT 26
 Q7RM79 PRELIMINARY; PRT; 779 AA.
 ID Q7RM79;
 AC Q7RM79;
 DT 01-MAR-2004 (TremBLrel. 26, Created)
 DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
 DE GTPase of unknown function, putative.
 GN Name=PY02306;
 OS Plasmodium yoelii yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=12368865;
 RA Carlton J.M., Anguioi S.V., Suh B.B., Koolij T.W., Perteza M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 RT parasite plasmodium yoelii yoelii";
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an

CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
 CC Bacillus cereus group; Bacillus anthracis.
 OX NCBI_TaxID=261594;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ames 0581;
 RX MEDLINE=99445483; PubMed=10515943;
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
 RT "Bacillus anthracis comparative genomics";
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE017336; AAT28865.2; -;
 DR EMBL; AE017336; AAT28865.2; -;
 DR PIR; B59102; B59102.
 DR TIGR; GBAA_px01_0124; -;
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF00395; SLH; 3.
 DR PROSITE; PS01072; SLH DOMAIN; UNKNOWN_1.
 KW Complete proteome; Plasmid.
 SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

 Query Match 7.7%; Score 116.5; DB 2; Length 652;
 Best Local Similarity 19.0%; Pred. No. 22;
 Matches 59; Conservative 65; Mismatches 106; Indels 81; Gaps 10;

 QY 25 DFYKYLQVWPWTFTETIKELSRFKQEQYSQASVL-----VGDIKVLMDSQDYFE 78
 DB 237 DVINQKIDE---FDKLSQRKDLERMLELNQKLSQLKQSPQLQDLNKLKESQSRLE 293

 QY 79 ATQ-----TVYEWGCVVTVLLSAYILLFDYNE--KASQAQKDLIRILDGYNK 126
 DB 294 LNKXDSNRLELNSEIKLNDRAKELLSLIMELIKQSEFDKIKNEKDDLNKREDLINR 353

 QY 127 LNEAQK-----SLGSSQSFNNASGKLALDSQLTNDSEKSSYFQ---SQV 170
 DB 354 IAESKELAKKKAELNTKLVELFKVQEAALNKKSGQYLYINKLDNELRELADKYKNSDNKI 413

 QY 171 DRIRKEAVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVK 230
 DB 414 SRLKNH-----IGEYNKQLEKIE-----ELE 435

 QY 231 QANKOIDAAKLKAT-ETAAIGEIKTETTFYVDYDDL-----MLSLKGAAKMINT 284
 DB 436 ECNKKIDNTTKQLAEFDKSNKKQOESELSVLQNKIDELGKRHKHQELESQSKALDE 495

 QY 285 CNEYQQRHGKK 295
 DB 496 AKEINKKLAEK 506

 RESULT 26
 Q7RM79 PRELIMINARY; PRT; 779 AA.
 ID Q7RM79;
 AC Q7RM79;
 DT 01-MAR-2004 (TremBLrel. 26, Created)
 DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
 DE GTPase of unknown function, putative.
 GN Name=PY02306;
 OS Plasmodium yoelii yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=12368865;
 RA Carlton J.M., Anguioi S.V., Suh B.B., Koolij T.W., Perteza M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 RT parasite plasmodium yoelii yoelii";
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an


```
QY 79 ATQTVYEWGVVTVQLLSAYILLDFEYNEKASAKDILIRILDGWNKLNKAEKSLGSS 138
Db 83 GKSVVEEAKKTAIEKLOAYLATNLAEEKLEAKPLIKVAD-----ATLERLASI 134
QY 139 OSFNNASGKLALDSQTLNDFSEKSSYFQSQVD-----RIRKEAY-----A 179
Db 135 LSAEDAALTAFTVSELVPAIDPVSAKPSLVDQKIAQEHSSGLYRASRTISLVA 194
QY 180 GAAAGIVAGPGLIISYSIAAGVIEGKLI-----PELNDRLKAVONF 221
Db 195 IIVGVLIAGTAGLITRISITGLAEGVEVANRLAAGDLTVFVAGRGRDETQOLMAAGNM 254
QY 222 FTSL-----SVTVKQANKOIDAAKLKLAT---EIAA-IGRIKETE 258
Db 255 VTSRLHLTAELAISISHGIASASNLHATSEQIATGSEVSAQVAVATASE 305

RESULT 29
Q81298
ID Q81298 PRELIMINARY; PRT; 941 AA.
AC Q81298; Q6I440; Q6KXV0;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2004 (TREMELrel. 28, Last annotation update)
DE Conserved domain protein.
GN OrderedLocNames=BA0374, BAS0360; ORFNAMES=GBAA0374;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R.J., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benson J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Sterne;
RA Bretin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017025; AAP24404.1; -
DR EMBL; AE017334; AAT29469.1; -
DR EMBL; AE017225; AAT52691.1; -
DR TIGR; BA0374; -
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR008994; Nucleic acid OB.
DR SEQUENCE 941 AA; 103896 MW; 78AF6B5381D7B6C6 CRC64;

Query Match 7.7%; Score 116; DB 2; Length 941;
Best Local Similarity 20.1%; Pred. No. 37;
```

```
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7;
QY 7 EQTVVVKSAIETADGAL-DFYNKYLDQVIP-WKTFDETIKELSRFKQYEQSEASVLVGD 64
Db 596 KEVVESINEATQNASAQLGDFLATYDSEIIVNFNTAERTKMSKNTSQILKEADKKLPD 655
QY 65 IKVLLMSQDKYFEATQTVYEWGVVTVQLLSAYILLDFEYNEKASAKDILIRILDG 124
Db 656 VKLLEDSK-----GLV-----DGR 671
QY 125 NKLENAQKSLGSSQSFNNASGKLALDSQ-----LTNDFSEKSSYFQSQVDRIK 175
Db 672 KKLADIKAEMPATEKKIKELADKIRDFESEEDLKDIIIRLLKNDVEKQSDYFANPNLKEN 731
QY 176 EAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELND---RLKAVQNFF---TSLSVT 228
Db 732 KLFAMPNPGSAMSPPYTVLWVGLLWVSLITVEVHEEGANYKSHIYFGRLLTFLTMG 791
QY 229 VKQA 232
Db 792 LSQA 795

RESULT 30
AAT29469
ID AAT29469 PRELIMINARY; PRT; 941 AA.
AC AAT29469;
DT 01-JUN-2004 (TREMELrel. 27, Created)
DT 01-JUN-2004 (TREMELrel. 27, Last sequence update)
DT 01-JUN-2004 (TREMELrel. 27, Last annotation update)
DE Conserved domain protein.
GN GBAA0374.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT29469.1; -
DR SEQUENCE 941 AA; 103896 MW; 78AF6B5381D7B6C6 CRC64;

Query Match 7.7%; Score 116; DB 2; Length 941;
Best Local Similarity 20.1%; Pred. No. 37;
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7;
QY 7 EQTVVVKSAIETADGAL-DFYNKYLDQVIP-WKTFDETIKELSRFKQYEQSEASVLVGD 64
Db 596 KEVVESINEATQNASAQLGDFLATYDSEIIVNFNTAERTKMSKNTSQILKEADKKLPD 655
QY 65 IKVLLMSQDKYFEATQTVYEWGVVTVQLLSAYILLDFEYNEKASAKDILIRILDG 124
Db 656 VKLLEDSK-----GLV-----DGR 671
QY 125 NKLENAQKSLGSSQSFNNASGKLALDSQ-----LTNDFSEKSSYFQSQVDRIK 175
Db 672 KKLADIKAEMPATEKKIKELADKIRDFESEEDLKDIIIRLLKNDVEKQSDYFANPNLKEN 731
QY 176 EAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELND---RLKAVQNFF---TSLSVT 228
Db 732 KLFAMPNPGSAMSPPYTVLWVGLLWVSLITVEVHEEGANYKSHIYFGRLLTFLTMG 791
QY 229 VKQA 232
Db 792 LSQA 795

RESULT 31
HLV1_ECOLI
```

HLX1_ECOLI STANDARD; PRT; 1023 AA.
 P09983;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Hemolysin, chromosomal.
 GN Name=hlyA;
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J96 / Serotype O4;
 RX MEDLINE=85234404; PubMed=3891743;
 RA Felmlee T., Pellett S., Welch R.A.;
 RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";
 RL J. Bacteriol. 163:94-105(1985).
 RN [2]
 RP SEQUENCE OF 1-44 FROM N.A.
 RC STRAIN=2001;
 RX MEDLINE=85258115; PubMed=3894051;
 RA Nicaud J.-M., Mackman N., Gray L., Holland I.B.;
 RT "Characterisation of HlyC and mechanism of activation and secretion of
 haemolysin from E. coli 2001.";
 RL FEBS Lett. 187:339-344(1985).
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 cell membranes and cause cell rupture by mechanisms not clearly
 defined.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DOMAIN: The Gly-rich region is probably involved in binding
 calcium, which is required for target cell-binding or cytolytic
 activity.
 CC -!- DOMAIN: The three transmembrane domains are believed to be
 involved in pore formation by the cytotoxin.
 CC -!- FTM: Palmitoylated by hlyC. The toxin only becomes active when
 modified.
 CC -!- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly
 by strains causing extraintestinal infections, such as those of
 the urinary tract.
 CC -!- SIMILARITY: Belongs to the RTX prokaryotic toxin family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M10133; AAA23975.1; -;
 CC EMBL; X02768; CAA26546.1; -;
 CC PIR; A24433; LEECA.
 CC InterPro; IPR001343; Hemlysn_Ca_bind.
 CC InterPro; IPR003995; RTXA.
 CC InterPro; IPR011049; Serralysn_like_C.
 CC Pfam; PF00353; HemolysinCbind; 6.
 CC Pfam; PF02382; RTX; 1.
 CC PRINTS; PRO00313; CAENDNGRPT.
 CC PRINTS; PRO1488; RTXTOXINA.
 CC PROSITE; PS00330; HEMOLYSIN_CALCIDIUM; 4.
 CC KEGG; Cytolysis; Hemolysis; Lipoprotein; Palmitate; Repeat; Toxin;
 KW Transmembrane.
 FT TRANSMEM 237 259 Potential.
 FT TRANSMEM 267 326 Potential.
 FT TRANSMEM 364 410 Potential.
 FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.
 FT REPEAT 723 728 1.
 FT REPEAT 732 737 2.
 FT REPEAT 741 746 3.
 FT REPEAT 750 755 4.
 FT REPEAT 759 764 5.
 FT REPEAT 768 773 6.

FT REPEAT 782 791 7.
 FT REPEAT 786 791 8.
 FT REPEAT 795 800 9.
 FT REPEAT 806 812 10.
 FT REPEAT 816 821 11.
 FT REPEAT 825 830 12.
 FT REPEAT 834 839 13.
 FT REPEAT 843 848 14.
 FT REPEAT 855 860 15.
 FT REPEAT 864 869 16.
 FT LIPID 563 563 N(6)-palmitoyl lysine (By similarity).
 FT LIPID 589 589 N(6)-palmitoyl lysine (By similarity).
 FT VARIANT 6 A -> T (in strain 2001).
 SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;
 Query Match 7.7%; Score 116; DB 1; Length 1023;
 Best Local Similarity 20.2%; Pred No. 41;
 Matches 68; Conservative 59; Mismatches 112; Indels 98; Gaps 11;
 QY 2 TGIFAEQTVVVK-----SAIETADGALDFVNYKLDQVPMKTFDTETIKELSRFKQE 53
 DB 155 TALSSMKIDELIKKQKSGNVSSSELSAKASTELINQLVDTA-----ASLNNVNSFSQ 207
 QY 54 YSOEASVL-----VGDIKVLMDSQDKFEATQTVYEWGCVVTVLLSAYILL-FDEY 104
 DB 208 LNKGLSVLSNTKHLNGVGN-KLQNLPLNDITGAGLDTV---SGILSAISASFILSNADAD 263
 QY 105 NEKSAQOKDILIRLDD---GVNKLNEAQKSLGSSOSFNNASGKLALDLSQLTNDFSE 161
 DB 264 TGTRAAAGVEITTKVLGVNGKISQYIIAQRRAQGLST----- 302
 QY 162 KSYFQSQVDRIKREAYAGAAGIVAGPFGLIIS----YSIAAGVIEGKLIPELNDRKA 217
 DB 303 -----AAAAGLIASVVTLLAISPLSLFIADKPKRANKIEYSQRPKK 344
 QY 218 VQNFSTLSVTVKQANKDIDAUKLATEIAAIGIKTETETTRFYVYDDMLSLKGA 277
 DB 345 LGYDGSLSLAAAFHKETGAIDASTRISTVLASVSGISAAATTSLV----GAPVSALVGA 400
 QY 278 -----AKKMINTCNEYOORHGK 294
 DB 401 VTGIISGILEASKQAMPEHVASKADVAEWEKKGK 437

RESULT 32

RA50_METJA STANDARD; PRT; 1005 AA.
 AC Q58718;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE DNA double-strand break repair rad50 ATPase.
 GN Name=rad50; OrderedLocNames=MJ1322;
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=868087;
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.W., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The

ID	Accession	Organism	Length	Score	Score 115	DB 2	Length 1093
Q928G7	PRELIMINARY;	PRT; 1093 AA.					
ID	Q928G7	01-DEC-2001 (TrEMBLrel. 19, Created)					
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)						
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)						
DE	Lin2568 protein.						
GN	OrderedLocusNames=lin2568;						
OS	Listeria innocua.						
OCC	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.						
NCBI_TaxID=1642;							
SEQUENCE FROM N.A.							
STRAIN=CLIP 11262 / Serovar 6a;							
MEDLINE=21537279; PubMed=11679669;							
Glaser P., Frangeul L., Buchrieser C., Rusnik C., Amend A.,							
Baquerot F., Berche P., Blocher H., Brandt P., Chakraborty T.,							
Charbit A., Chetoui F., Couve E., de Daruvar A., Deboux P.,							
Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,							
Entian K.-D., Feihi H., Garcia-del Portillo F., Garrido P.,							
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,							
Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,							
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,							
Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,							
Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,							
Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;							
"Comparative genomics of <i>Listeria</i> species."							
Science 294:849-852(2001).							
EMBL; AL596172; CAC97795.1; ..							
PIR; AC1753; AC1753.							
ListList; LIN2568; ..							
InterPro; IPR010090; Tape_meas_TP901.							
InterPro; IPR007713; TMP.							
Pfam; PF05017; TMP; 11.							
TIGRFAMs; TIGR01760; tape_meas_TP901; 1.							
Complete proteome.							
SEQUENCE 1093 AA; 121152 MW; 67AAA59058B9B4D4 CRC64;							
Query Match	7.6%;	Score 115;	DB 2;	Length 1093;			
Best Local Similarity	22.6%;	Pred. No. 52;	Mismatches 127;	Indels 76;	Gaps 11;		
Matches 71;	Conservative 40;						
QY	2	TCGFAEQTVVEVVKSAIETADGALDFYKYL-----					31
Db	270	TKGFNKNFGLVKGEAKTKTIVDFYNTGLVDSYEAGEALTQTKQLNDLNSTDLSEVT					329
QY	32	DOVIPW-KTFD-----ETIKELSRPKQBSYQEAFLVGDIVKLLMSQDKYFEATQTVIEW					86
Db	330	EKALAFSKTFDSMDNETLGRNALMETYGLSARQSFDMTVGAQGNLKNKTDLGDLNLAAY					389
QY	87	CGVVTQLLSAYLLFD-----EYN-EKASAKDILIRLLDDGVKNLNEAOKSLIG					136
Db	390	SQSPQKNGYSQAQGFTELAGLDSGAYNLKDNVLKFGIRVSDSDIEKAVEG-----LG					445
QY	137	SSQSPFNNAAGKLLALDSOLTNDTFSEKSYFQSQVDRIKREAVAGAAGIVAGPFLI---					193
Db	446	-----GKWKTYADMKDKGYDNNELFANLATEINKVGDQERASIVSAIFGSLGSD					496
QY	194	--ISYSIAAGVIEGKLIPELNDRLKAVQNFTSL---SVTVKANKDIDAACL--KLATE					246
Db	497	NTVKVLTAMG-----DLNGELGAVQGGYDDVKGASEKLITETNSKQDLTKMHWELQTA					548
QY	247	IAAIGEIKTETETT 260					
Db	549	LAPIGEYLLLELANT 562					
RESULT 34							
Q08581	PRELIMINARY;	PRT; 821 AA.					
ID	Q08581	01-NOV-1996 (TrEMBLrel. 01, Created)					
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)						
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)						


```
DR PROSITE: PS00422; GRANINS_1; UNKNOWN_1.
KW Hypothetical protein. 176615 MW; 9AE7AB0AABF6EAE CRC64;
SQ SEQUENCE 1478 AA; 7.5%; Score 113.5; DB 2; Length 1478;
Query Match
Best Local Similarity 20.9%; Pred. No. 94;
Matches 68; Conservative 51; Mismatches 98; Indels 109; Gaps 15;
QY 5 FAEQTVVVKSAIETADGALDFYNYK-----DQVWPKTDFE-TIKELSRFKQ 52
DB 328 FFEKLTLEKEKEEYRNELDNKLLHAEQNKKKLDQLERYKNDDEHIKSL----- 382
QY 53 EYSQASVLVGDIKVLLMDSQDKYFEATQTVYVGVVTVLLSAVILLDFSYNEKASQAQ 112
DB 383 ---KESEIEINEKNTLILEQKLAQAS---YE-----ISMEKSNKKN--- 422
QY 113 KDILRIILDDGVN---KLNKAEQKSLGSSQSFNNASGKLLALDSQUTNDFSEKSYFQSQV 170
DB 423 -----NNSYNGERNSEYEKKL---BELNN-----ITNSYEKEINELNKEK 459
QY 171 DRIRK---EAVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLPELND-----RLKAVQNF 221
DB 460 ENITQFMKIMDDDEGI-----NKLKEELNDANALIVNLKNKNE 500
QY 222 FTSLSVTVKQANKDI-DAAKLKLATEIAAIGRIKTKETETTFYVDYDDLMLSLKGAAXK 280
DB 501 LYNVNIEMEQANKWRDDIDILLAN-----IDKINDERKVTENEKEQKELKYNELKINYE 556
QY 281 MINTCN-----EYQQRH 292
DB 557 KYKECNKFFNPLPKMKKKIEYEKKH 582
RESULT 37
RECN AQUAE STANDARD; PRT; 520 AA.
AC O66834;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE DNA repair protein recN (Recombination protein N).
GN Name=recN; OrderedLocusNames=AQ_561;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: May be involved in recombinational repair of damaged DNA
CC (By similarity).
CC -!- SIMILARITY: Belongs to the recN family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; A800695; AAC06789.1; -.
DR PIR; F70350; F70350.
DR HSSP; P25054; 1DEB.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004604; RecN.
DR InterPro; IPR003405; SMC_C.
```

```
DR Pfam; PF02483; SMC_C; 1.
DR TIGRFAMs; TIGR00634; recN; 1.
KW ATP-binding; Complete proteome; DNA repair.
FT NP_BIND 29 36 ATP (Potential).
SQ SEQUENCE 520 AA; 60439 MW; A9708562ACBE901E CRC64;
Query Match
Best Local Similarity 7.5%; Score 112.5; DB 1; Length 520;
Matches 59; Conservative 70; Mismatches 94; Indels 71; Gaps 13;
QY 10 VEVVKSIAETADGALDFYNYKLDQVWPKTDFE-----TIKELSRFKQYSQ---EASVL- 61
DB 135 LEKVYSLRKKKEQLFELRKKKEELIKQDYLERVREVEEIGISSEYELKKNKANLIN 194
QY 62 -VGDIKVLMLDSQDKYFEATQTVYVGVVTVLLSAVILLDFSYNEKASQAQDI-- 115
DB 195 NLEKVKKAVGESLYKLLGENSVYIIEIRKLNKAKVESYSGKFSLEIKIANLEEEVYE 254
QY 116 -----LIRILDGVNKLNEAOKSLGSSQSFNNASGKLL-----ALDSQLTN---DF 159
DB 255 LYNLSKEEMPEISEEVNEINEKLFRIQRLKEKYKSPPEILKEVEEIKELSINLVDF 314
QY 160 SEKSSYFQSQVDRIKRAYAGAAAGIVAGPFGLLIISYSIAAGVIEGKLPELNDRLKAV 218
DB 315 KEEE--LREEVEKLREE-----YDKLAEEVSRRDKKA 345
QY 219 QNFFTSLSVTVKQANKDI-DAAKLKLATEIAAIGRIKTKETETTFYVDYDDLMLSL 272
DB 346 EDLEERIEILKEUN--LERAKLV-----EIK-ESEPTKYGKDKIEFLFS 388
RESULT 38
SVR SYN3
ID SYR SYN3 STANDARD; PRT; 584 AA.
AC Q55486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN Name=argS; OrderedLocusNames=all0502;
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; D64006; BAA10833.1; -.
DR PIR; S75986; S75986.
DR HSSP; Q93RP5; 1IQO.
DR HAMAP; MF_00123; -.
DR InterPro; IPR005148; ArgtrNASynthet_N.
DR InterPro; IPR001278; Arg tRNA-synt 1c.
DR InterPro; IPR008909; tRNA-synt_1d_C.
```



```

DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR009080; tRNA-synt 1a bind.
DR Pfam; PF03485; Arg tRNA synt N; 1.
DR Pfam; PF00750; tRNA-synt 1d; 1.
DR Pfam; PF05746; tRNA-synt 1d C; 1.
DR PRINTS; PR01038; tRNA-syntHARG.
DR TIGRfam; TIGR00456; argS; 1.
DR PROSITE; PS00178; AA tRNA ligase I; 1.
KW Aminoacyl-tRNA synthetase; ATP-binding; Complete proteome; Ligase;
KW Protein biosynthesis.
FT SITE 126 136 "HIGH" region.
SQ SEQUENCE 584 AA; 65212 MW; 89602A1A5A3BD85B CRC64;

Query Match 7.5%; Score 112.5; DB 1; Length 584;
Best Local Similarity 23.2%; Pred. No. 36;
Matches 76; Conservative 45; Mismatches 131; Indels 75; Gaps 16;

QY 3 GIFAQTVVVKSAIETADGALD-----FY---NKYLDVPIPKTFDETIKELSRFKQE 53
DB 171 GMLITLKEVPEALVTAD-ALDIGLVTFYKQAKQRFQD-----DEQFRETSR----- 218
QY 54 YSQEASVLVGDIK-----VLLMDSQDKYPEATQTVYEWGVVQ-----LLSAYIL 99
DB 219 -QAVVALQAGDAKSIKAWLLCEQSRREF---QLIYDCLDITIERGSEFYNPFLPGVVE 274
QY 100 LFDEYNEKKAQKDILIR-----ILDDGVNKLNEAQSLLGSSQ-SFNNAQKLLAL 151
DB 275 LLQE-----KOLLVEDNGAQCFLDGTNKGDRPLIVKQSDGGYNVATTDLAAL 325
QY 152 DSQLTNDPSEKSYF--QSQVDRIRKEAYAGAAAGIVAGPFLGIISYIAAGVIEGKLIP 209
DB 326 NYRLNTDGAKEIIYVTDAGQANHFQAQFFQVAKAGILTDPTQV---HVPFLVKGEDGK 382
QY 210 ELNDRLKAVONPFTSLSVTVKQAKDID-----AAKLKATRIA---AIGETK---T 255
DB 393 KLKTRAGDTIRLKKLLTEAVTARQDLTRLTAEERSTEEFKTEVAVRGVIGAVKYADL 442
QY 256 ETETRFYVDYDDLMLSLKGAAKMI 282
DB 443 SQNRTSDYVFPDKMLAQNGTAPYML 469

RESULT 39
Q6CGC3 PRELIMINARY; PRT; 941 AA.
AC Q6CGC3
DT 01-OCT-2004 (TREMELrel. 28, Created)
DT 01-OCT-2004 (TREMELrel. 28, Last sequence update)
DE Similar to tr|Q8WZS2 Neurospora crassa Dynactin.
GN ORFNames=VALIOA20504g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVRES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Bolesane A., Boyer J., Catillico C., Confaiolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dunazet H., Groppi A.,
RA Hanttraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennens D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;

"Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RC SEQUENCE FROM N.A.
RA STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382127; CAG84227.1; -.
SQ SEQUENCE 941 AA; 106891 MW; EF36A7BBE6D40AFB CRC64;

Query Match 7.5%; Score 112.5; DB 2; Length 941;
Best Local Similarity 23.4%; Pred. No. 84;
Matches 78; Conservative 49; Mismatches 118; Indels 89; Gaps 15;

QY 6 AEQTVVVKSAIET-ADGALDFYINKYLDQVWP-----KTFDETIKELSRFKQESQAS 59
DB 547 SEQLEQSVNSYIDTVSEGAL-----LVDEAVNVPVFQTFDDFLSTQTDYNTYLDMSK 601
QY 60 VLVGD-IKVILM-----DSQK-----YPEATQTVYEWGVV 91
DB 602 LAIHTRVSVILLKQLGHRSCANSLSKLYWIDNQDKTARVPSPIIDNAVQTLSDVSFAM 661
QY 92 QLL-----SAVILLFDEYNEKKA-OKDILIRLLDDGVNKLNEAOKSLGSSQSFNA 144
DB 662 QLLHDDDPDSAYTAIFKRSSKRPJSAVEKDI-----KLTE-QLSTLDFSVSQND 711
QY 145 SGKLLALDSQLTNDPSEKSYFQSQVDRIRKEAYAGAAAGIVAGPFLGIISYIAAGVIE 204
DB 712 ---LPTWAQLOPQSQESIVDTISADVHLKAVL-----LKKETITETLEIK 755
QY 205 GKLIPELMDRLKAVONPFTSL-----SVTVKQAKDIDAAKLKATEIAAIGIKET 257
DB 756 IKHYDSKAERPAEKEEAFVQLRANLNEAALTEKRLRDDVESLIKAKLEQKNADARFKTHV 815
QY 258 ETRFYVDYDDLMLSLKLG--AAKGMINTCEYQ 289
DB 816 Q-----NRKLSTLKGEDIVDMHLNTHQSQ 841

RESULT 40
HLVA_ECOLI
ID HLVA_ECOLI STANDARD; PRT; 1024 AA.
AC P08715;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin, plasmid.
GN Name=hlyA;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Hesse J., Wells W., Vogel M., Goebel W.;
RT "Nucleotide sequence of a plasmid-encoded hemolysin determinant and
RT its comparison with a corresponding chromosomal hemolysin sequence.";
RL PEMS Microbiol. Lett. 34:1-11(1986).
RN [2]
RP PALMITOYLATION OF LYS-564 AND LYS-690.
RX MEDLINE=9509325; PubMed=7801126;
RA Stanley P., Packman L.C., Koronakis V., Hughes C.;
RT "Fatty acylation of two internal lysine residues required for the
RT toxic activity of Escherichia coli hemolysin.";
RL Science 266:1992-1996(1994).
RN [3]
RP PALMITOYLATION OF LYS-564 AND LYS-690.
RX MEDLINE=96404790; PubMed=8808931;
RA Ludwig A., Garcia P., Bauer S., Jarchau T., Benz R., Hoppe J.,
RA Goebel W.;
RT "Analysis of the in vivo activation of hemolysin (HlyA) from
RT Escherichia coli.";
```

RL J. Bacteriol. 178:5422-5430(1996).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- SURCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: The gly-rich region is probably involved in binding
CC calcium, which is required for target cell-binding or cytolytic
CC activity.
CC -!- DOMAIN: The three transmembrane domains are believed to be
CC involved in pore formation by the cytotoxin.
CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when
CC modified.
CC -!- MISCELLANEOUS: The hemolysin of E. coli is produced predominantly
CC by strains causing extraintestinal infections, such as those of
CC the urinary tract.
CC -!- SIMILARITY: Belongs to the RTX prokaryotic toxin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14107; AAA98233.1; --
CC InterPro; IPR001343; Hemlysn_Ca_bind.
CC InterPro; IPR003995; RTX.
CC InterPro; IPR011049; Serralyen_like_C.
CC Pfam; PF00353; HemolysinCbind; 6.
CC Pfam; PF02382; RTX; 1.
CC PRINTS; PR00313; CABNDNGRPT.
CC PRINTS; PR01488; RTXTOXINA.
CC PROSITE; PS00330; HEMOLYSIN_CALCIIUM; 4.
CC Calcium; Cytolysis; Hemolysis; Lipoprotein; Palmitate; Plasmid;
CC Repeat; Toxin; Transmembrane.
CC TRANSMEM 238 260 Potential.
CC TRANSMEM 268 327 Potential.
CC TRANSMEM 365 411 Potential.
CC DOMAIN 724 870 16 X REPEATS, GLY-RICH.
CC REPEAT 724 729 1.
CC REPEAT 733 738 2.
CC REPEAT 742 747 3.
CC REPEAT 751 756 4.
CC REPEAT 760 765 5.
CC REPEAT 769 774 6.
CC REPEAT 778 783 7.
CC REPEAT 787 792 8.
CC REPEAT 796 801 9.
CC REPEAT 807 812 10.
CC REPEAT 817 822 11.
CC REPEAT 826 831 12.
CC REPEAT 835 840 13.
CC REPEAT 844 849 14.
CC REPEAT 856 861 15.
CC REPEAT 865 870 16.
CC LIPID 564 564 N(6)-palmitoyl lysine.
CC LIPID 690 690 N(6)-palmitoyl lysine.
CC SEQUENCE 1024 AA; 110201 MW; 83944917F76C945B CRC64;

Query Match 7.5%; Score 112.5; DB 1; Length 1024;
Best Local Similarity 20.5%; Pred. No. 71;
Matches 69; Conservative 55; Mismatches 116; Indels 97; Gaps 11;

QY 2 TGFAEQTVVVK-----SAIETADGALDFNVKLVQVDPKTFDETIKELSRKQE 53
DB 155 TALSSMKIDELIKQKSGNGVSSSELAQASIELINQLVDTVASLN-----NNVNSFSQ 208
QY 54 YSQEASVY-----VGDIKVLMDSDQKYFEATQTVYEWCGVTVQLLSAYILL-FDEY 104
DB 209 LNTLGSVLSTKHLNGVGN-KLQNLPLNDIGAGLDTV---SGILSAIGASFILSNADAD 264
QY 105 NEKASAKQDILIRLDD-----GVNKLNEAQSLILGSSQSFNNASGKLLALDLSQLTDFSE 161

Db 265 TRTKAAAGVELTTKVLGNVGKGISQYIIAQRAAQLST----- 303
QY 162 KSSYFOSQVDRTRKEAYAGAGAGIVAGPGLIIS-----YSIAAGVIEGKLIPELNDRLKA 217
Db 304 -----AAAAGLIASAVTLAISPLSFLSIADKFRANKIEYSORPKK 345
QY 218 VQNFFTSLSVTVKQANKDIDAAKLKIATRIATGAEIKTETETTRFYVDYDDMLSLKGA 277
Db 346 LGYDGSLLAAAFHKETGAIDASLTISTVLASVSSGISAATTSLV---GAPVSALVGA 401
QY 278 -----AKMINTCNEYQORHCK 294
Db 402 VTGIISGLSEASQAMFEHVASQADVIABWEKKHCK 438

RESULT 41
OS0733 PRELIMINARY; PRT; 1098 AA.
AC OS0733;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Hypothetical protein BBG10.
GN OrderedLocustNames=BBG10;
OS Borrelia burgdorferi (lyme disease spirochete).
OC Plasmid lp28-2.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OC NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/375551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lachiga R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwin M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Matthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi";
RL Nature 390:580-586(1997).
DR EMBL; AB000786; AAC66075.1; --
DR PIR; B70232; B70232.
DR TIGR; BBG10; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR Pfam; PF00015; MCPsignal; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 1098 AA; 123861 MW; F465D18421F05935 CRC64;

Query Match 7.5%; Score 112.5; DB 2; Length 1098;
Best Local Similarity 20.6%; Pred. No. 77;
Matches 79; Conservative 60; Mismatches 122; Indels 123; Gaps 17;

QY 10 VEVVKSALJETADGALDFNVKY-----LDQVIMPKTFDETIKELSR 49
Db 552 VQVEKSVNFEFNSNVDFVNEYNQLLKEKESREREIITLPTDQVSALQKLNDEINEKNK 611
QY 50 -FKQYSQEASVLDGIK--VLLMDSQKYFEATQ-----TVYEWCGV 90
Db 612 AFVEYKGSFETLNEESNRQVVVALEKQVNEKTDALDRSFVEAKALQKEITDLEW---E 668
QY 91 TOLLISAYILLFDEYNEKASAKQDILIRI-----LDD---GVNKLNEAQK- 132
Db 669 TMLLPA-----KERASAEKQWASKIQAQMYKFFVDEHKSQPKLNETNRTIKQYAE 719
QY 133 -----SLLGSSQSFNNA-----SGKLLALDS-----QLTNDSEKSSYFSQV 170

```

Db 720 KAQDTTYSYDSMIDGLNFQNAFMKDIAGKFLNKDTGESIGEPFNLINGKDVNMGEG 779
QY 171 DRIKREAYAG-----AAAGIVAGPGLIISYSIAA-----GVIRGKLIPDLNRLKAV 218
Db 780 EXWTTQMYESWKTGLTAGAVFGPGEAVBELINGLTDVFWGILKG-----GEKARIKAI 835
QY 219 QNFFTSLSVTVKQAKNDIAAKLKLATEIAAI-----GKIKTETETFRFYVDYDDMLSL 273
Db 836 E-----KKRDBLSELEKRSVELKLEKLEDRPDEBIKMKRKEKLSLDDEYTKIEIF 885
QY 274 LKGAKKMINTCNEVQOR-HGKKT 296
Db 886 LKQAQSGKQISGEEFQKRLHVDQT 909

RESULT 42
ID Q6C359 PRELIMINARY; PRT; 1906 AA.
AC Q6C359;
DT 01-OCT-2004 (TREMBLrel. 28, Created)
DT 01-OCT-2004 (TREMBLrel. 28, Last sequence update)
DE 01-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Similar to DEHAOC09658g Debaryomyces hansenii IPF 1836.1.
GN ORFNames=YALIOF02387g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]_TaxID=4952;
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anhouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckrich J.M., Beyne E., Bleykasten C.,
RA Boibrame A., Boyer J., Cattolico L., Confantolero F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Leaur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Winkler P., Souciet J.L.;
FT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382132; CAG77705.1; -.
SQ SEQUENCE 1906 AA; 210346 MW; F6ED7D1AF7B7562B CRC64;

Query Match
Best Local Similarity 22.1%; Pred. No. 1.5e+02;
Matches 66; Conservative 49; Mismatches 115; Indels 69; Gaps 12;

QY 7 EOTVEVKSATAGALDFYNYL-----DQVIPWKTPTDETIKELSRFKQYSQBSVLIV 62
Db 1195 EKELETKTSELTAKAELTKKELTAKSDEA---TTYSAKVKELETSAALEKQTTLK 1251
QY 63 GDKVLLMDSQKYPEATVTVWCGVTVQLLSAYILLFDYENKASAKQDILIRLDD 122
Db 1252 AMADNLTKDLAEK-----TKELVAAKSEL---ESSNTSSKEEDVLTAKGLSD 1295
QY 123 GYNKLENAQKSLGSSQSFNNASGKLLALDLSLTNDPSEKSSYFOSQVDRIKAEVAGAA 182
Db 1296 ATAEEVELKKS---SQAAETEASSKVSLEAKLTK-ASESS---KAELEDKVNK----- 1341

```

```

QY 183 AGIVAGPGLIISYSIAAGVIE-----GKLIPELNDRLKAVQNF---FTLSVTVKQA 232
Db 1342 -----LSSFKELQTSKDDHSTSVKSLTEQVRESTLKAENFEHDISSLKDDLAQA 1392
QY 233 NKDIDAAKLKLATEIAAAGEIKTETETFRFYVDYDDMLSLKGAKKMINTCNEVQOR 291
Db 1393 EKERDALRTELDTSIK-----EMENERTSLTKDAD-----SATKELTNKVMLOTK 1438

RESULT 43
O45614 PRELIMINARY; PRT; 3102 AA.
AC O45614; P91824; Q9TZR4;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE C. elegans LAM-3 protein (Corresponding sequence T22A3.8) (Laminin
DE alpha chain) (Hypothetical protein T22A3.8).
GN Name=lam-3; Synonyms=lamal/2; ORFNames=T22A3.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Barlow K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=22668894; PubMed=12783803;
RA Huang C.C., Hall D.H., Hedgecock E.M., Kao G., Karantz V.,
RA Vogel B.E., Hutter H., Chisholm A.D., Yurchenco P.D., Wadsworth W.G.;
RT "Laminin alpha subunits and their role in C. elegans development.";
RL Development 130:3343-3358 (2003).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.A.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL008585; CAA15432.3; -.
DR EMBL; Z81125; CAA15432.3; JOINED.
DR EMBL; AL008585; CAB03385.3; JOINED.
DR EMBL; Z81125; CAB03385.3; -.
DR EMBL; AF074902; AAC26793.1; -.
DR FIR; F87908; F87908.
DR FIR; T23064; T23064.
DR PIR; T43291; T43291.
DR HSSP; P02468; INPE.
DR WormPep; T22A3.8; CE31067.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0005606; C:laminin-1; IEA.
DR GO; GO:0005103; F:receptor binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030155; P:regulation of cell adhesion; IEA.
DR GO; GO:0030334; P:regulation of cell migration; IEA.
DR GO; GO:0045995; P:regulation of embryonic development; IEA.
DR InterPro; IPR008985; Cona_like_lec_g1.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR009030; Grow fac recept.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.

```

InterPro; IPR009254; Laminin I.
 DR InterPro; IPR010307; Laminin II.
 DR InterPro; IPR008211; Laminin N.
 DR InterPro; IPR003129; TSP N.
 DR Pfam; PF00052; Laminin B; 2.
 DR Pfam; PF00053; Laminin_EGF; 17.
 DR Pfam; PF02210; Laminin_C2; 3.
 DR Pfam; PF06008; Laminin I; 1.
 DR Pfam; PF06009; Laminin II; 1.
 DR Pfam; PF06055; Laminin N; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRODOM; PD003031; Laminin B; 2.
 DR SMART; SM00180; EGF_Lam; 17.
 DR SMART; SM00281; LamB; 2.
 DR SMART; SM00282; LamG; 4.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF 1; 12.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
 DR PROSITE; PS00025; LAM G_DOMAIN; 3.
 KW Hypothetical protein; Laminin EGF-like domain.
 SQ SEQUENCE 3102 AA; 343790 MW; 2EE9305D5C591FAB CRC64;

Query Match 7.5%; Score 112.5; DB 2; Length 3102;
 Best Local Similarity 19.9%; Pred. No. 2.6e+02;
 Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

QY 6 ABEQTVIEW-----KSAIETADGALDFYNNKLDQVPIPKWTPDETTELKLSRFKQYEQAS 59
 DB 1862 ABEELKQWTAQKLNITFEDLNKRDVLEQWMD-----YRETIYDVSKDADARMS 1915

QY 60 VLVG-----DIKVLMDSDQK-----YFEATQTV----- 83
 DB 1916 LVVGKRNRYKEVSEIEKRLVEAEQIAYSRSNIEKARSELMWFEKKNLTLAEL 1975

QY 84 ---YEWCGVVTQLLSAYILLFFDEYNEK-----KASAQKDILIRILDGWNKLN 128
 DB 1976 PDLVEQCNITLL---YSQLIDYEDYEVYVOTAGRAHAEKLEVOAQK-IVDRFVDVTRTETEN 2031

QY 129 -----BAQKS-----LLGSSQSFNNAGSKLLALDSQLTNDPSEKSSYFQSQVDRI 173
 DB 2032 PLKASHAYENIVEALKNATEAVDSAAEASVSKMLGSGSGGDANEES--LRSLQELK 2089

QY 174 RKEAYAGAAAGIVAGPFLGLIISYIAAGVIE--GKLIPELNDRLKAVQNFFTSLSVTVKQ 231
 DB 2090 KNESLSN-----VNSNAVKIVEELKKEKDLTRLGHNLKTSI----- 2131

QY 232 ANKDDIAAKLKLATEIAAIGEIKETETTRFYVDYDDMLSLKGAQKQ-INTCNEYQQ 290
 DB 2132 -----VKRLGVKNKNEASS---WDDKDRMHSLKNGAKTAHERSANVCKE 2173

QY 291 RHGKKT 296
 DB 2174 SEGIKT 2179

RESULT 44
 Q6GNE7 PRELIMINARY; PRT; 1012 AA.
 ID Q6GNE7
 AC Q6GNE7;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RX MEDLINE=22388257; PubMed=12477932;

Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Schenken C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 FAhey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 EN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RA Klein S., Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC073565; AAH73565.1; -
 DR InterPro; IPR009053; Prefoldin.
 DR InterPro; IPR007794; Rib_recept_KP.
 DR Pfam; PF05104; Rib_recpt_KP_reg; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1012 AA; 112965 MW; CF23AB8E0D9EDAEEA CRC64;

Query Match 7.4%; Score 112; DB 2; Length 1012;
 Best Local Similarity 20.7%; Pred. No. 75;
 Matches 68; Conservative 61; Mismatches 127; Indels 72; Gaps 14;

QY 8 QTVFVKSATFADGALDFYNNKLD-QVWP-----KTFDETIKLSRFKQYEQ 56
 DB 665 QIASIQKEAELKEAVEQCKNNKNDLREKNQWQAMEALGLAEKTCCKLNSEKAKBEMVQ 724

QY 57 EASVLVGDIKVLMD-----SQDYFRATQTVWCGVVTQLLSAYILLFFDEYNEKK 108
 DB 725 QLSAVQSQTKETLOSIVLPQITIVSQSYSEWLQ---EPRRTSQQLS-----QOTEKE 774

QY 109 ASAQKDILIRILDGWNKLN---BAQKSLLGSSQSFNNAGSKLLALDSQLTNDPSEKSSY 165
 DB 775 GSSELQQLQKQEDSQSDLOVECEKYRTILQTEAM-----LKALQNSVEESEEQVWKAK 828

QY 166 FQSQVDRIRKEAYAGAAAGIVAGPFLGLIISYIAAGVIEGKLIPELNDRLKAVQNFFTSL 225
 DB 829 FSSSEELR-----SHSQVKTLEE-----TVCKLRDLQSTQOLKECV 867

QY 226 SVTVKQANKDIDAAKILKAT---ETAAIGEIKETET-----ETTRFYVDYDDMLSLKGA 278
 DB 868 SLMEAQLETQWNAKTECQTVSNELISQQLSSEHRLDATKAEARQSIELSIVLRQOL 927

QY 279 KKMINTCNEYQQRHG---KKTL--LEVP 301
 DB 928 GEMLNHVND-TEKHGTETHTQIQNEVP 954

RESULT 45
 Q7RFQ7 PRELIMINARY; PRT; 1081 AA.
 ID Q7RFQ7

Cell 42:173-182(1985).
-!- FUNCTION: Not known but may be related to activation of the
variant surface glycoprotein genes.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M11452; AAA30191.1; -
DR PIR; A03395; VMUT21.
DR InterPro; IPR004922; ESAG.
DR Pfam; PF03238; ESAG1; 1.
KW Signal.
FT SIGNAL 1 23
FT CHAIN 24 329 VSG expression site-associated protein
221A.
FT CARBOHYD 73 73 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 294 294 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 308 308 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;

Query Match 7.4%; Score 111; DB 1; Length 329;
Best Local Similarity 22.9%; Pred. No. 23;
Matches 48; Conservative 33; Mismatches 63; Indels 66; Gaps 9;

QY 92 QLLSAVILLFD-----EYNEKKASAKQDILIRILDGYNKL-NEAQS 133
Db 4 EIVELVLLFSVTCVDAMWLOGADCTRVADHKEHAPVTEAVCYLRCLSDALKLYSEGEKK 63
QY 134 LIGSSQSFNNAGKLLALDSQTNDFSEKSY-----FQSQVDRIRKEAYAGAAIV 186
Db 64 LLVTEEVYANAS---LILD-DMEGRAGESSTYLSVIRGVMEQTDRLK----- 108
QY 187 AGPFGLLISYSIAAGVI---EGKLIPELNDRLKAVQ-----NFTSLSVTVKQA 232
Db 109 -----LISYGNMGNLVAKAGLFALEDLSKEVRKEIPGALIKTKNYTSVAEIVRTV 162
QY 233 NKDIDAANKL-----ATEIAAIGEIKTE 256
Db 163 WEDVGEILWKETEAKCGSKQVGVGEIQTE 192

RESULT 48
Q6ZM87 PRELIMINARY; PRT; 329 AA.
AC Q6ZM87
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ESAG1.
GN Names=H25N7.33; Synonyms=ESAG1;
OS Trypanosoma brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=98434256; PubMed=9763292;
RA Rudenko G., Chaves I., Dirks-Mulder A., Borst P.;
RT "Selection for activation of a new variant surface glycoprotein gene
expression site in Trypanosoma brucei can result in deletion of the
old one.";
RT old one.";
RL EMBL; AL671259; CAD21890.1; -
RN [2]
RP SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;
RC STRAIN=427;
RX MEDLINE=22103446; PubMed=12106867;
RA Berriman M., Hall N., Shneider K., Bringaud F., Tiwari B., Isobe T.,
Bowman S., Corton C., Clark L., Cross G.A.M., Hoek M., Zanders T.,

Barberof M., Borst P., Rudenko G.;
RT "The architecture of variant surface glycoprotein gene expression
sites in Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 122:131-140(2002).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=427;
RC STRAIN=427;
RA Zeng C., Zhao B., Hiehl M., Catanese J., Gertard C., Melville S.E.,
Hoek M., Navarro M., Cross G.A.M., El-Sayed N., Barberof M.,
Rudenko G., Borst P., de Jong P.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL671259; CAD21890.1; -
DR InterPro; IPR004922; ESAG.
DR Pfam; PF03238; ESAG1; 1.
SQ SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;

Query Match 7.4%; Score 111; DB 2; Length 329;
Best Local Similarity 22.9%; Pred. No. 23;
Matches 48; Conservative 33; Mismatches 63; Indels 66; Gaps 9;

QY 92 QLLSAVILLFD-----EYNEKKASAKQDILIRILDGYNKL-NEAQS 133
Db 4 EIVELVLLFSVTCVDAMWLOGADCTRVADHKEHAPVTEAVCYLRCLSDALKLYSEGEKK 63
QY 134 LIGSSQSFNNAGKLLALDSQTNDFSEKSY-----FQSQVDRIRKEAYAGAAIV 186
Db 64 LLVTEEVYANAS---LILD-DMEGRAGESSTYLSVIRGVMEQTDRLK----- 108
QY 187 AGPFGLLISYSIAAGVI---EGKLIPELNDRLKAVQ-----NFTSLSVTVKQA 232
Db 109 -----LISYGNMGNLVAKAGLFALEDLSKEVRKEIPGALIKTKNYTSVAEIVRTV 162
QY 233 NKDIDAANKL-----ATEIAAIGEIKTE 256
Db 163 WEDVGEILWKETEAKCGSKQVGVGEIQTE 192

RESULT 49
CAD21890 PRELIMINARY; PRT; 329 AA.
AC CAD21890;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE ESAG1.
GN H25N7.33.
OS Trypanosoma brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=98434256; PubMed=9763292;
RA Berriman M., Hall N., Shneider K., Bringaud F., Tiwari B., Isobe T.,
Bowman S., Corton C., Clark L., Cross G.A.M., Hoek M., Zanders T.,
Barberof M., Borst P., Rudenko G.;
RT "The architecture of variant surface glycoprotein gene expression
sites in Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 122:131-140(2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=98434256; PubMed=9763292;
RA Rudenko G., Chaves I., Dirks-Mulder A., Borst P.;
RT "Selection for activation of a new variant surface glycoprotein gene
expression site in trypanosoma brucei can result in deletion of the
old one.";
RT old one.";
RL EMBL; AL671259; CAD21890.1; -
RN [2]
RP SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEE8 CRC64;
RC STRAIN=427;
RX MEDLINE=22103446; PubMed=12106867;
RA Berriman M., Hall N., Shneider K., Bringaud F., Tiwari B., Isobe T.,
Bowman S., Corton C., Clark L., Cross G.A.M., Hoek M., Zanders T.,

QY 92 QLLSAVILLFD-----EYNEKSAQKDIILIRILDDGVNKL-NEAOKS 133
DB 4 EIVELVLLFVSVTCVDAWLOGADCTRVADHKEHAPVTEAVCYLRCLSDALNKLXSEGEKK 63
QY 134 LIGSSQSFNNASGKLLALDSLTNDFSEKSSY-----FOSQVDRIRKEAVAGAAAGIV 186
DB 64 LLVTEEVANAS---LIID-DWEGRAGESSTVLSVIRGMEQTDRLK----- 108
QY 187 AGPFLIISYSIAAGVI---EGKLIPELNDRLKAVQ-----NFFTSLSVTVKQA 232
DB 109 -----LISYGNKGMNLVAKAGGLFALEDSLKEVRKEIPGALIKNTKYTTSVAEIVRTV 162
QY 233 NKDIDAAXKL-----ATEIAAIGEIKTE 256
DB 163 WEDVGEILWKETEAKCGSQKVEGVGEIQTE 192

RESULT 50

Q73AJ3 PRELIMINARY; PRT; 465 AA.
AC Q73AJ3;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Flagellin.
GN OrderedLocusNames=BCE1780;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pXOI.;"
RL Nucleic Acids Res. 32:977-988(2004).
DR EMBL; AE017269; AAS40708.1; -;
DR TIGR; BCE1780; -;
DR InterPro; IPR001029; Flagellin_C.
DR InterPro; IPR001492; Flagellin_N.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 1.
KW Complete proteome; Flagellum.
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

Query Match 7.4%; Score 111; DB 2; Length 465;
Best Local Similarity 22.0%; Pred. No. 35;
Matches 70; Conservative 51; Mismatches 129; Indels 68; Gaps 14;
QY 6 AEQTEVVVKSAIETADGALDFYNNKYLDQV---IPWKTDETIKELSRPK-----Q 52
DB 184 AKATI-TAKDAPDAKADSDAAKEDIAAAKDTPSKNDQAQLEAKYIEAKATLATLKPTDA 242
QY 53 EYSQEAASVL-----VGDIKLLMDSDQKYPEATQVYEWCGVVTQLLSAVILLFDEYNE 106
DB 243 TYAKAAELDRAATNALNDNAKNVLDVGYEKKLTITTKKA-----EYTA 285
QY 107 KKASAQKDIILIRILDDGVNKLNEAQKSLGSSQSFNNASGKLLALDSLTNDFSEKSSYP 166
DB 286 AKEQSTKSTAAA---DLVTKYETAKSNALGNDIAKEYLEAK-TAYEAN-KNDISSKSR-F 339
QY 167 QSQVDRIRKEAVAGAAAGIVAGPGLIISYSTA--AGVIEGKL--IPELNDRLKAVQNPF 222
DB 340 EAAETELNKDITANKAAKY-----LVETYEKATGATGTEKSLVAVDKIDEALKTIADNR 393
QY 223 TSLSVTVKQAKNDIAAKKLKATEIAAIGEI-----KTETETTRFYVDYDDLMLSLK 276
DB 394 ATLGATLRLDPNVNVLKQSASMASAASQIEDADMAKEMSEWTKF-----KILNE 444

QY 277 AAKQMINTCNEYQQRHGK 294
DB 445 AGISMLSQANQTPOQVSK 462

Search completed: January 28, 2005, 19:16:04
Job time : 205 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:16:10 ; Search time 143 Seconds

(without alignments)
765.529 Million cell updates/sec

Title: US-09-993-292B-24

Perfect score: 1510

Sequence: 1 MTGIPAEQTVVKSIAETA.....TCNEYQQRHGKKTLLVPDI 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1466	97.1	305	9	US-09-993-292A-2
2	136.5	9.0	889	16	US-10-437-963-197045
3	116	7.7	891	14	US-10-369-493-10918
4	116	7.7	941	15	US-10-282-122A-46343
5	116	7.7	1023	10	US-09-884-696-5
6	115	7.6	1005	14	US-10-369-493-1061
7	112.5	7.5	2823	14	US-10-369-493-5220
8	112.5	7.5	2823	14	US-10-369-493-5221
9	111	7.4	1295	14	US-10-369-493-6440
10	110.5	7.3	564	15	US-10-424-599-212070
11	110	7.3	1189	15	US-10-282-122A-70920
12	108.5	7.2	1959	14	US-10-028-248A-106
13	108.5	7.2	1959	15	US-10-107-782-106

14	108.5	7.2	1999	14	US-10-028-248A-107	Sequence 107, App
15	108.5	7.2	1999	15	US-10-107-782-107	Sequence 107, App
16	108	7.2	815	16	US-10-437-963-191043	Sequence 191043, A
17	108	7.2	1965	14	US-10-369-493-3279	Sequence 3279, App
18	107.5	7.1	742	15	US-10-282-122A-53523	Sequence 53523, A
19	107.5	7.1	857	14	US-10-369-493-9374	Sequence 9374, App
20	106.5	7.1	1938	14	US-10-369-493-5109	Sequence 5109, App
21	106.5	7.1	1938	14	US-10-369-493-5110	Sequence 5110, App
22	106.5	7.1	3692	15	US-10-282-122A-71235	Sequence 71235, A
23	106	7.0	1961	14	US-10-028-248A-105	Sequence 105, App
24	106	7.0	1961	15	US-10-107-782-105	Sequence 105, App
25	105.5	7.0	344	14	US-10-369-493-497	Sequence 497, App
26	105	7.0	737	16	US-10-437-963-107270	Sequence 107270, A
27	104	6.9	722	9	US-09-815-242-10796	Sequence 10796, A
28	103.5	6.9	660	10	US-09-841-260-139	Sequence 139, App
29	103.5	6.9	660	13	US-10-007-693-139	Sequence 139, App
30	103.5	6.9	660	16	US-10-762-058-139	Sequence 139, App
31	103.5	6.9	1020	15	US-10-282-122A-58016	Sequence 58016, A
32	103.5	6.9	1163	15	US-10-282-122A-51864	Sequence 51864, A
33	103.5	6.9	5373	16	US-10-408-765A-741	Sequence 741, App
34	103	6.8	862	14	US-10-369-493-19762	Sequence 19762, A
35	103	6.8	961	14	US-10-080-608A-66	Sequence 66, Appl
36	103	6.8	961	14	US-10-370-685-155	Sequence 155, App
37	103	6.8	1583	15	US-10-408-765A-1635	Sequence 1635, App
38	102.5	6.8	472	15	US-10-282-122A-55283	Sequence 5283, A
39	102	6.8	590	15	US-10-220-480-50	Sequence 50, Appl
40	102	6.8	590	16	US-10-220-481-155	Sequence 155, App
41	102	6.8	612	15	US-10-220-480-56	Sequence 56, Appl
42	102	6.8	612	15	US-10-220-481-161	Sequence 161, App
43	102	6.8	1903	17	US-10-766-993-3	Sequence 3, Appl
44	102	6.8	1938	14	US-10-032-585-7611	Sequence 7611, App
45	101.5	6.7	1948	14	US-10-171-311-164	Sequence 164, App
46	101.5	6.7	1945	10	US-09-927-597-2	Sequence 2, Appl
47	101.5	6.7	1972	14	US-10-171-311-162	Sequence 162, App
48	101.5	6.7	1972	14	US-10-341-434-103	Sequence 103, App
49	101.5	6.7	1979	10	US-09-927-597-4	Sequence 4, Appl
50	101	6.7	1475	14	US-10-369-493-6362	Sequence 6362, App
51	100.5	6.7	425	14	US-10-434-418-3	Sequence 3, Appl
52	100.5	6.7	526	14	US-10-308-936-15	Sequence 15, Appl
53	100.5	6.7	529	15	US-10-282-122A-48338	Sequence 48338, A
54	100.5	6.7	553	14	US-10-308-936-14	Sequence 14, Appl
55	100.5	6.7	609	14	US-10-308-936-12	Sequence 12, Appl
56	100.5	6.7	609	16	US-10-408-765A-566	Sequence 566, App
57	100.5	6.7	609	16	US-10-408-765A-2133	Sequence 2133, App
58	100.5	6.7	613	14	US-10-013-477-18	Sequence 18, Appl
59	100.5	6.7	613	14	US-10-013-477-26	Sequence 26, Appl
60	100.5	6.7	613	14	US-10-207-655-204	Sequence 204, App
61	100.5	6.7	613	14	US-10-308-936-9	Sequence 9, Appl
62	100.5	6.7	613	15	US-10-284-237-2002	Sequence 2002, App
63	100.5	6.7	613	16	US-10-408-765A-571	Sequence 571, App
64	100.5	6.7	613	17	US-10-770-668-10	Sequence 10, Appl
65	100.5	6.7	623	14	US-10-308-936-11	Sequence 11, Appl
66	100.5	6.7	627	14	US-10-308-936-8	Sequence 8, Appl
67	100.5	6.7	803	9	US-09-815-242-5597	Sequence 5597, App
68	100.5	6.7	805	9	US-09-815-242-12286	Sequence 12286, A
69	100.5	6.7	861	14	US-10-310-154-485	Sequence 485, App
70	100.5	6.7	861	14	US-10-369-493-17514	Sequence 17514, A
71	100.5	6.7	1009	15	US-10-282-122A-43788	Sequence 43788, A
72	100	6.6	451	14	US-10-369-493-2102	Sequence 2102, App
73	100	6.6	523	10	US-09-976-782-6	Sequence 6, Appl
74	100	6.6	680	15	US-10-425-114-68086	Sequence 68086, A
75	100	6.6	957	16	US-10-437-963-176713	Sequence 176713, A
76	100	6.6	990	15	US-10-108-260A-3896	Sequence 3896, App
77	100	6.6	1174	14	US-10-094-749-1911	Sequence 1911, App
78	100	6.6	1939	17	US-10-798-037-4	Sequence 4, Appl
79	100	6.6	2481	15	US-10-467-595-23	Sequence 23, Appl
80	99.5	6.6	857	14	US-10-369-493-9166	Sequence 9166, App
81	99.5	6.6	1959	14	US-10-028-248A-36	Sequence 36, Appl
82	99.5	6.6	1959	15	US-10-107-782-36	Sequence 36, Appl
83	99.5	6.6	1961	14	US-10-028-248A-103	Sequence 103, App
84	99.5	6.6	1961	15	US-10-107-782-103	Sequence 103, App
85	99	6.6	268	9	US-09-815-242-5337	Sequence 5337, App
86	99	6.6	520	14	US-10-369-493-1809	Sequence 1809, App

RESULT 3
US-10-369-493-10918
; Sequence 10918, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10918
; LENGTH: 891

```
; TYPE: PRT
; ORGANISM: Ferroplasma acidarmanus
US-10-369-493-10918

Query Match
Best Local Similarity 7.7%; Score 116; DB 14; Length 891;
Matches 69; Conservative 46; Mismatches 129; Indels 66; Gaps 12;

Qy 4 IFPAETVWVKSATADGALDFYNNKYLDQVI-----PWKTFDETIKELS 48
Db IFSQEIETAAASYINEIK-SLETEKKELENTIENASKGHELFNVLSAENENLEKAIKELQ 563
Qy 49 RPKQYISQASVLDGIVKVLMD--SQKYFEAT-QTVYEWGVVTVQLLSAYILLDFBYN 105
Db QYNEYIYRYSIIISG-----MDPGAIIKEIEAAGESFTTFKRNINELLS-QIGFVPEYN 616
Qy 106 EKASAKDILIRILDGDKVNLNEAKSILGSSQSFNNASGKLLALDSQLTNDPSEKSSY 165
Db EVONTGKISSEINRLKTEVERSRMKSLESIKDIEINRKSISIGRIEMENKQSAMHQY 676
Qy 166 FQSQVDRIKAYAGAAAGVAGPPFGLIISYSIAAGVIEGKLIPELNDRLK---AVQNFF 222
Db -----DGIDEQ-----AGNIESRYKSAYQENIKMKTIVDSYT 708
Qy 223 TSLSVTVKQA-NKQIDAAKLKATE-IAAIGHEIKTETETTPYVDYDMLSLKGAACK 280
Db ERIEETENAKNLEQDAEKYKKTREAITLGLK-----REAFDNGIQSIIRKDSAS 761
Qy 281 MINTCNEYQQ 290
Db MTNLTRKYLQ 771

RESULT 4
US-10-282-122A-46343
; Sequence 46343, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46343
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46343

Query Match
Best Local Similarity 7.7%; Score 116; DB 15; Length 941;
Matches 49; Conservative 45; Mismatches 88; Indels 62; Gaps 7;

Qy 7 EQTVVWVKSATETADGAL-DFYNNKYLDQVIP-WKTFDETIKELSRFKQYEQEASVLVGD 64
Db KEVVESINEATQNASAQLGDFLATYDSEIVNFENFATERTKMSKNTSQILKEADKKLPD 655
Qy 65 IKVLLMSQDKYFEATQTVYEWGVVTVQLLSAYILLDFEYNEKKASAKODILIRILDGV 124
Db VKLLEDSK-----GLV-----DGR 671
Qy 125 NKLEAOKSLGSSQSFNNASGKLLALDSQ-----LTNDFSEKSSYFQSOVDRIK 175
Db KGLADIKAEMPATEKKIKELADKIRDFESEBDLKIIRLLKNDVEKQSDYFANPNVKEN 731
Qy 176 BAYAGAAGIVAGPFGLIISYSIAAGVIEGKLIPELND---RLKAVQNFF---TSLSVT 228
Db KLFAMPNVGSAWSPFYTVLAWGALLMVSLLTVEVHEGANYKSHIEYFGRLLTLFTMG 791
Qy 229 VKQA 232
Db LSQA 795

RESULT 5
US-09-884-696-5
; Sequence 5, Application US/09884696
; Publication No. US20030035809A1
; GENERAL INFORMATION:
; APPLICANT: GEORGE, LISLE W
; APPLICANT: ANGELOS, JOHN A
; APPLICANT: HESS, JOHN F
; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES
; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA
; TITLE OF INVENTION: BOVIS INFECTIONS
; FILE REFERENCE: 481.06
; CURRENT APPLICATION NUMBER: US/09/884,696
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-884-696-5

Query Match
Best Local Similarity 7.7%; Score 116; DB 10; Length 1023;
Matches 68; Conservative 59; Mismatches 112; Indels 98; Gaps 11;

Qy 2 TGIFAEQTVVVK-----SAIETADGALDFYNNKYLDQVIPWKTFTETIKELSRFKQE 53
Db TALSSMKIDELIKKQSGNVSSSELAKASTELINQLVDTA-----ASLNNVNSFSQQ 207
Qy 54 YSQEASVL-----VGDIKVLLMSQDKYFEATQTVYEWGVVTVQLLSAYILL-PDEY 104
Db LNKLGSVLSNTKHLNGVGN-KLQNLPLNDNTGAGLDTV---SGILSAISASFILSNAD 263
Qy 105 NEKKASAKODILIRILD---GVNKLNEAOKSLGSSQSFNNASGKLLALDSQLTNDPSE 161
Db TGTAAAGVELTTKVLGVNGKISQYIIAQRAAOGLSTS-----302
```

162	QY	KSSYFQSQVDRIRKEAYAGAAAGIVAGPFGIIIS----	YSIAAGVIEGKLIPELNDRLKA	217
303	Db	-----AAAAGLIASVVTLAISPLSFLSIADKFRANKRIBEYSORFKK	344	
218	QY	VQNFFTSLSVTVKQANKDIDAAKKLATAEIAIGIKTETTFVYDDDLMSLLKA	277	
345	Db	LYGDGDSLLAAFHKETGAIDASLSTRISVLASVSSGISAAATTSLV----	400	
278	QY	-----AKQINTCNEYQOEHGK	294	
401	Db	VTGIISGILEASKOAMFEHVASKADYIAWSEKKHGK	437	

```

RESULT 6
US-10-369-493-1061
; Sequence 1061, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1061
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-10-369-493-1061

```

```

Query Match      7.6%; Score 115; DB 14; Length 1005;+
Best Local Similarity 19.4%; Pred. No. 0.9;
Matches 59; Conservative 54; Mismatches 93; Indels 98; Gaps 11;

QY      7 EQTVVWVKSATETADGALDFPNKYLDQVIEPWKTFDE-----TIKELSRF-----KQEYSQ 56
DB      605 DEILEDIKSQLNFKF---NFTNQVLASVSYLNSVDEGINRRIKELTENIVSGWNKEKRE 661
QY      57 EASVL-----VGDIKVLLMDSODKYFEATQTVVEMCGVVTQLLSAYILLFPEYNKEKAS 110
DB      662 ELNKLREDEREINRLKDKLNLKNKEKE-----LIEIENRRSLKFKDYKEYILGL 710
QY      111 AQKDILIRILDGVNKNLNEAKSLGSSQSPNNASGKLIALD----- 152
DB      711 TEKLEELKNKIDGLSEI-----YNICNSKILAIKDINKKYKNKEDIEIYLNKK 757
QY      153 -----SOLTDPFEKSSYFQSQVDRIKREAYAGAAAGIVAGPFGLLIISYTAAGVIEGKLI 208
DB      758 ILEVNKKEINDIEERISYINQKLUDE-----INYNEE-----EHKKI 792
QY      209 PEL-----NDRLKAVONFFTSLSVTVQKANDIDAAKLKLATEIAAIGEIKTETETTRFVY 264
DB      793 KELYENKRQELDNVREQTEIETIGIEYLKQDVESIKARL-----KEMSNLEKEKELTRFV 848
QY      265 DYDD 268
DB      849 EYLD 852

```

RESULT 7
US-10-369-493-5220
; Sequence 5220, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

```

1  APPLICANT: Slater, Steven C.
2  APPLICANT: Goldman, Barry S.
3  APPLICANT: Chen, Xiaofeng
4  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
5  TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
6  FILE REFERENCE: 38-10(52052)B
7  CURRENT APPLICATION NUMBER: US/10/369,493
8  CURRENT FILING DATE: 2003-02-28
9  PRIOR APPLICATION NUMBER: US 60/360,039
10 PRIOR FILING DATE: 2002-02-21
11 NUMBER OF SEQ ID NOS: 47374
12 SEQ ID NO 5220
13 LENGTH: 2823
14 TYPE: PRT
15 ORGANISM: Caenorhabditis elegans
16 10-369-493-5220

```

[illegible]

RESULT 8
US-10-369-493-5221
Sequence 5221, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 5221
LENGTH: 2823
TYPE: PRT
ORGANISM: *Caenorhabditis elegans*
US-10-369-493-5221

Query Match 7.5%; Score 112.5; DB 14; Length 2823;
Best Local Similarity 19.9%; Pred. No. 6;
Matches 73; Conservative 64; Mismatches 106; Indels 123; Gaps 16;

Qy 6 ABOQTEVV-----KSAIETADGALDFYKYLQVDPKTKETIKELSRFKQBYSQEAS 59
Db 1862 ABEILKMVTAQKJNETIFEDLKNRIDVLEQWMD-----YRETIYDVSKKDTADAERMS 1915
Qy 60 VLVG-----DIKVLMDSDQK-----YFEATQTV----- 83
Db 1916 LVVGRINRYKEVNEIEKLREVAEDQIAYSNSIEKARSELMNMFEDKEKINMTLAE 1975
Qy 84 ---YEWCGVVTQLLSAYILLFDEYNEK-----KASAQKDLIRILDGYNKLN 128
Db 1976 PDLVEQCNITLL---YSLIDEYDEEYVQTAGRHAKELEVOAQK-IVDRFVDTRTETEN 2031
Qy 129 -----EAKS-----LLGSSQSFNNASGKLLALDSQTNDFSEKSSYFQSDVRI 173
Db 2032 PLKASHAYENIVEALKNATEAVDSAAEASAVSKMLGSESGDANBES---LRSQLEKL 2089
Qy 174 RKEAYAGAAAGIVAGPFGLLIISYSIAAGVIE--GKLIPELNDRLKAVQNFPTSLSVTVKQ 231
Db 2090 KNESSLSN-----VDNSNAVKIVELKEKEDLDRGLHNLKTSI----- 2131
Qy 232 ANKOIDAAKLAATEIAAIGBIKTETETTRFYVDYDDLMLSLKGAAKKM-INTCNEYQQ 290
Db 2132 -----VKRLGVKNEASS---WDDKHDRMHSILKNGAKTAKHSAERNSANVCKE 2173
Qy 291 RHGKKT 296
Db 2174 SEGINT 2179

RESULT 9
US-10-369-493-6440
; Sequence 6440, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6440
; LENGTH: 1295
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6440

Query Match 7.4%; Score 111; DB 14; Length 1295;
Best Local Similarity 21.0%; Pred. No. 2.8;
Matches 65; Conservative 58; Mismatches 93; Indels 94; Gaps 13;

Qy 6 ABOQTEVVKSAIETADGALDFYKYLQVDPKTKFDET-----IKELSRFKQBYSQEASVL 61
Db 710 ABETNEKLRSLASSE-----EQILDKNQOESLIDDLKEKLHSAESTQELQVS 759
Qy 62 VGDIKVLMDSDQKQFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASA-----QKDI 115
Db 760 LEMLKIEVSNAQKQWSE-----VLKESFEALQLELSAQSVRSRVVDAVQEKDG 811
Qy 116 LIRILDGYNKLENAQS-----LLGSS-----QSFN---NASCK- 147
Db 812 LLRLVDTLKLKIEDTEKSAQDLQSSVBEIKQLDLQNFQKNAFVLESLEKLNSSHR 871

Qy 148 LIALDSQTNDFSEKSSYFQSQVDRIKREAYAGAAAGIVAGPFGLLIISYSIAAGVIEGK 206
Db 872 DMVALASQL-EELQHLVWGESQVENVKEEL-----IGAKIMNKE 910
Qy 207 LIPELN---DRLKAVQNFPTSLSVT-----VKQANKDIDAAKLKLATEIAAIG 251
Db 911 MVDELNAKLGDALLEGMBELKKSEVSEAKVQRREBELIAQVSKHRDQOE-QLQTLDELK 969
Qy 252 EIKTETETTR 261
Db 970 SAQSTETSTR 979

RESULT 10
US-10-424-599-212070
; Sequence 212070, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 212070
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Glycine max
; FEATISM:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_33525C.1.pep
US-10-424-599-212070

Query Match 7.3%; Score 110.5; DB 15; Length 564;
Best Local Similarity 18.9%; Pred. No. 0.99;
Matches 57; Conservative 59; Mismatches 115; Indels 71; Gaps 9;

Qy 11 EVKSAIETADGALDFYKYLQVDPKTKFDETIKELSRFKQBYSQEASVLVGP---IKV 67
Db 61 EVEREKLRVAEANLE-----KQAMDWMLAQEELKRLGEDAARHABESSETLDFRRVKK 114
Qy 68 LLMDSDQKQFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASQA-----KDILIRILD 121
Db 115 LLNDVRSSELVSSQALASSRSKMEEQERLLELQSELGEQASVMSYMNKDAQIEV-- 172
Qy 122 DGVNKLNEAQSLLGSSQSFNNASGKLLALDSQTNDFSEKSSYFQSQVDRIKREAYAGA 181
Db 173 -----ESERTKLRVAESRRNRELELDLMEKELISELEELKERTSLEQAVKE----- 220
Qy 182 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNDRLKAVQNFPTSLSVTVKQANKDIDA 241
Db 221 -----VALQSELEKKTAEFRETSAVLQVKESELVD 254
Qy 242 KLATEIAAIGBIKTETETTRFYVDYDDLMLSLKGAAKKMINTCNE---YQOR---HGKKT 296
Db 255 E-----IQELKSEKASLQGLEEKLELS-----SARKMLGDVYNQEIYDLKMLMHSKET 303
Qy 297 LL 298
Db 304 QL 305

RESULT 11
US-10-282-122A-70920
; Sequence 70920, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos

Wed Feb 2 11:26:41 2005

APPLICANT: Malone, Cheryl
APPLICANT: Haseelbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patentin version 3.1
SEQ ID NO 70920
LENGTH: 1189
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70920
Query Match 7.3%; Score 110; DB 15; Length 1189;
Best Local Similarity 18.3%; Pred. No. 3;
Matches 61; Conservative 60; Mismatches 130; Indels 82; Gaps 9;
7 EQTVGVVK-----SALATDAGALDFYKYLDOVIPKTFDETIKELSRFKQYSEASV 60
163 EESAGVLKYKKAESIQKLDHTEDNLNRVEDILYDLGRVEPLKEEAATAKEYKQLSKE 222
61 LVGDIKVLMDSDQKYFEATQTVYEWG-----GVVTTLLSAYILLFDEYNE 106
223 MEQSDVIVTSDIDHYTDNQRLDERLNHLKSOAEKEGQQAQINQLQRY----- 273
107 KKAQAQKDLIRILDDGVNKLNEAQSILGASSOSFNNAAGKLLALDSQTLND-----FS 160
274 -KGRQQN-----DYDIEKLN---YELVKATENVEQLSGKLNVLBERKKNQSETNARVE 323
161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLLIYSIAAGVIEGKLIPELNDL-----K 216
324 EELDNLSEQSDISIKNEKAQNE-----KLLADLKNKQKQNLN 359
217 AVQNFSTLSVTVKQANDIDAALKLATEIAAIGEIKTETETTRFYVDYDDLMLSLKG 276
360 EVOELESLLYVSDQDEKLEIKNSYTYLMSQSDVNDIRFLEHTINENEAQRSRLDS 419
277 AAKKMINTCNEYQQ-----RHGKKTLLV 300
420 RLVEAFNLQKDIQQNITQTNKAYQSSKKSMOKV 452
US-10-028-248A-106
Sequence 106, Application US/10028248A
Publication No. US20030235882A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard
APPLICANT: Patturajan, Meera
APPLICANT: Vernet, Corine
APPLICANT: Casman, Stacie
APPLICANT: Malyankar, Uriel
APPLICANT: Shenoy, Suresh
APPLICANT: Spytek, Kimberly
APPLICANT: Gangolli, Esha
APPLICANT: Miller, Charles
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Kekuda, Ramesh
APPLICANT: Smithson, Glemnda
APPLICANT: Zerhusen, Bryan
APPLICANT: Liu, Xiaohong
APPLICANT: Colman, Steven
APPLICANT: Tchernev, Velizar
APPLICANT: Si, Jingsheng
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Sciore, Paul
APPLICANT: Millet, Isabelle
APPLICANT: Rothenberg, Mark
TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
TITLE OF INVENTION: Thereof
FILE REFERENCE: 21402-222
CURRENT APPLICATION NUMBER: US/10/028,248A
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/256619
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 60/262959
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/272408
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/285189
PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: 60/308039
PRIOR FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 60/311266
PRIOR FILING DATE: 2001-08-09
NUMBER OF SEQ ID NOS: 211
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 106
LENGTH: 1959
TYPE: PRT
ORGANISM: Gallus gallus
US-10-028-248A-106
Query Match 7.2%; Score 108.5; DB 14; Length 1959;
Best Local Similarity 21.0%; Pred. No. 8.1;
Matches 66; Conservative 43; Mismatches 107; Indels 99; Gaps 11;
7 EQTVGVKSAE---TADGALDFYKYLDOVIPKTFDETIKE-----LSRFKQEYS 55
1166 EQEVTVLKKTLEDEAKTHEAOIQEMRQKHSQAI---EELAEQLEQTKRVKANLEKAKQALE 1223
56 QASVLVGDIDKVLMDSDQKYFEATQTVYEWGVTQLLSAYILLFDEYNEKASACKDI 115
1224 SERAELSENVKVLQKGD-----EHRKKVQDAQLOE 1256
116 L-----IRILDDGVNKLNEAQSILGASSOSFNNAAGKLLALDS---QLTNDFSEKSSY 165
1257 LQVKFTEGVRKTELAERVNKLQVEL-----DNVTGLNQSDSKSIKAKDFSALSEQ 1309
166 FQSQVDRIRKEAYAGAAAGIVAGPFGLLIYSIAAGVIEGK---LIPENLRLKAVQNF 222
1310 LQDTQELQEEET-----RLKLSFSTKLQTEDEKNALKEQLEEEEAARNLE 1356

ORGANISM: Rattus norvegicus
US-10-028-248A-107
Query Match
Best Local Similarity 7.2%; Score 108.5; DB 14; Length 1999;
Matches 66; Conservative 53; Mismatches 124; Indels 95; Gaps 13;
QY 7 EQTVVVKSAIE---TADGALDFYNYKLDQVIP--WKTFTET---IKELSRFKQYEQE 57
Db 1165 EQEVNLIKKTLEEEAKTHEAQIQEMRQKHSQAVEELAEQLEQTKRKVANLEKAKOTLENE 1224
QY 58 ASVLVGDIKVLLMSODKYFEATQTVYWCVVTTQLLSAYILLFDEYNEKKASAKDILI 117
Db 1225 RGLANEVKVLLQGGRD-----SEHKRKVEAQLOELQ 1257
QY 118 RILDGVNKLNEAQKSLGSSQSPNNAAGKLLALD---SOLTNDFS----- 160
Db 1258 VKFNEGERRVTELADKVTKLOVELDNVTGLSQSDSKSKLTQDFSALESQLODQTOELLQ 1317
QY 161 -----EKSYFQSVQDRIRKEAYAGAAAGIVAGPGLIISYSIAAGV 202
Db 1318 EENRQKLSLSTKLQVEDEKNS-FREQLEEEEBEAKHNLEKQIA-----TLHAQV 1366
QY 203 IEGKLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAACKLKLATEIAAIGBIKTETETRF 262
Db 1367 ADMK--KKMEDSVGCLL-----TAEVVKRKLQKDLGSLSORHEEKVAAYD--KLEKTKTRL 1418
QY 263 YVDYDDLMLSLKGAAKKMTNCTN-EYQQRHGKKTILE 299
Db 1419 QOELDDLLVDL-----DHQRSACNLEKKQKQKFDQLLAE 1452
RESULT 16
US-10-437-782-107
Sequence 107, Application US/10107782
Publication No. US20040018970A1
GENERAL INFORMATION:
APPLICANT: Boldog, Ferenc,
APPLICANT: Casman, Stacie,
APPLICANT: Colman, Steve,
APPLICANT: Edinger, Shlomit,
APPLICANT: Gangolli, Esha,
APPLICANT: Kekuda, Rameesh,
APPLICANT: Li, Li,
APPLICANT: Liu, Xiaohong,
APPLICANT: Malyankar, Uriel,
APPLICANT: Miller, Charles,
APPLICANT: Millet, Isabelle,
APPLICANT: Patturajan, Meera,
APPLICANT: Rothenberg, Mark,
APPLICANT: Sciore, Paul,
APPLICANT: Shenoy, Suresh,
APPLICANT: Shimkets, Richard,
APPLICANT: Si, Jingsheng,
APPLICANT: Smithson, Glennda,
APPLICANT: Spvtek, Kimberly,
APPLICANT: Stone, David,
APPLICANT: Taupier, Raymond, jr.,
APPLICANT: Tchernev, Velizar,
APPLICANT: Vernet, Corine,
APPLICANT: Zethusen, Brian
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
FILE REFERENCE: 21402-222CIP
CURRENT APPLICATION NUMBER: US/10/107,782
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: 10/028,248
PRIOR FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/256,619
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 60/262,959
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/272,408
PRIOR FILING DATE: 2001-02-28

Query Match
Best Local Similarity 7.2%; Score 108.5; DB 15; Length 1999;
Matches 66; Conservative 53; Mismatches 124; Indels 95; Gaps 13;
QY 7 EQTVVVKSAIE---TADGALDFYNYKLDQVIP--WKTFTET---IKELSRFKQYEQE 57
Db 1165 EQEVNLIKKTLEEEAKTHEAQIQEMRQKHSQAVEELAEQLEQTKRKVANLEKAKOTLENE 1224
QY 58 ASVLVGDIKVLLMSODKYFEATQTVYWCVVTTQLLSAYILLFDEYNEKKASAKDILI 117
Db 1225 RGLANEVKVLLQGGRD-----SEHKRKVEAQLOELQ 1257
QY 118 RILDGVNKLNEAQKSLGSSQSPNNAAGKLLALD---SOLTNDFS----- 160
Db 1258 VKFNEGERRVTELADKVTKLOVELDNVTGLSQSDSKSKLTQDFSALESQLODQTOELLQ 1317
QY 161 -----EKSYFQSVQDRIRKEAYAGAAAGIVAGPGLIISYSIAAGV 202
Db 1318 EENRQKLSLSTKLQVEDEKNS-FREQLEEEEBEAKHNLEKQIA-----TLHAQV 1366
QY 203 IEGKLIPELNDRLKAVQNFFTSLSVTVKQANKDIDAACKLKLATEIAAIGBIKTETETRF 262
Db 1367 ADMK--KKMEDSVGCLL-----TAEVVKRKLQKDLGSLSORHEEKVAAYD--KLEKTKTRL 1418
QY 263 YVDYDDLMLSLKGAAKKMTNCTN-EYQQRHGKKTILE 299
Db 1419 QOELDDLLVDL-----DHQRSACNLEKKQKQKFDQLLAE 1452
RESULT 15
US-10-107-782-107
Sequence 107, Application US/10107782
Publication No. US20040018970A1
GENERAL INFORMATION:
APPLICANT: Boldog, Ferenc,
APPLICANT: Casman, Stacie,
APPLICANT: Colman, Steve,
APPLICANT: Edinger, Shlomit,
APPLICANT: Gangolli, Esha,
APPLICANT: Kekuda, Rameesh,
APPLICANT: Li, Li,
APPLICANT: Liu, Xiaohong,
APPLICANT: Malyankar, Uriel,
APPLICANT: Miller, Charles,
APPLICANT: Millet, Isabelle,
APPLICANT: Patturajan, Meera,
APPLICANT: Rothenberg, Mark,
APPLICANT: Sciore, Paul,
APPLICANT: Shenoy, Suresh,
APPLICANT: Shimkets, Richard,
APPLICANT: Si, Jingsheng,
APPLICANT: Smithson, Glennda,
APPLICANT: Spvtek, Kimberly,
APPLICANT: Stone, David,
APPLICANT: Taupier, Raymond, jr.,
APPLICANT: Tchernev, Velizar,
APPLICANT: Vernet, Corine,
APPLICANT: Zethusen, Brian
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
FILE REFERENCE: 21402-222CIP
CURRENT APPLICATION NUMBER: US/10/107,782
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: 10/028,248
PRIOR FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/256,619
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 60/262,959
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/272,408
PRIOR FILING DATE: 2001-02-28

Best Local Similarity 20.7%; Pred. No. 2.7;
Matches 62; Conservative 62; Mismatches 95; Indels 80; Gaps 14;

QY 38 KTFDETIK-ELSRFKQEQYSSQ--EASVL-----VGDIKVLLMDSQDKYFEATQTVVEW 86
DB 230 KTLAEELREIRLESSLAQAGDDKDVLEAKLEKLGDNVIL-----QEKVLLSQEI-DN 344
QY 87 CGV-----VTQLLSA-----YILLDFEYNKKASQAQ-KDILIRILDGVNKNKNEAQKSLGS 137
DB 345 KGIRIRELSLLSSKEADYRNLCSFDQTKESLELAIAKIQQLEBEVHTRNDLSSKISS 404
QY 138 -----SOSFNASGKLLALDSQTNDFSEKSSYFQSQVDRIKREAYAGAAAGIVAGPF 190
DB 405 IDLNEELQALNSAKNEAEKLSLTDYDTOLKASSEARENSR-----449
QY 191 GLIISYSIAAGVIEGKLIPELNDRLKAVQNFTSLSVTKQANKDIDAACKLKLATEIAAI 250
DB 450 -LLEKONWIKQLGKLSALSSDKREN-----IAALKELDATKAMLENEVAV 500
QY 251 GBK-----TETETFRFYVDYD-----DLMLSLKGAAGKAMINTCNEYQQ 290
DB 501 KSLRESLQSTEALDTSRSEVSKLSVELDEANRNMQDLVLI-----SKLQDEFNMQE 554

RESULT 17
US-10-369-493-3279
; Sequence 3279, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10369, 493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3279
; LENGTH: 1965
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3279

Query Match 7.2%; Score 108; DB 14; Length 1965;
Best Local Similarity 21.3%; Pred. No. 8.9;
Matches 64; Conservative 47; Mismatches 113; Indels 76; Gaps 12;

QY 6 AEQTEVVVKSATETADGALDFYKYLDOVIPKTFD-----ETIKELSRFKQEQY 56
DB 57 AKYREALKQOELEHAKT-----NDWLDELTKNAEATKFRKEKGARAEQLRNMEDANS 112
QY 57 EASVLVGDIKVL---LMDSDQKYFEATQTVVEWC-----GVVTQLLSA---YILLFDE 103
DB 113 TIESITREQVLRKLEQAQDAEETLTKVQQLQEAARTTEGFKQELSAKRLVELKQ 172
QY 104 YNEKASAKQTLIR---ILDPGVNKNKNEAQKSLGSQSFNNSGKLLALDSQTNDFS 160
DB 173 QSETHRNRLKEVELRLEQIKDDHANEIRIRRELEQKE-----DHAQT---E 217
QY 161 EKSSYFQSQVDRIKE-----AYAGAAAGIVAGPGLIISISIAAG 201
DB 218 QBAQELQNEVDRIKASSDLGRSPGTGPQPKGDSFAATRAG---SPFGTPLS-----268
QY 202 VIEGKLIPELNDRLKAVQNFTSLSVTKQANK---DIDAACKLKLATEIAAIGETKETE 258
DB 269 -IRGRAGQATDALBELYNVKQLAGEKRCCKLQEBELDDAVAMLEAKMPEIDELNASE 327

RESULT 18
US-10-282-122A-53523
; Sequence 53523, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53523
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-282-122A-53523

Query Match 7.1%; Score 107.5; DB 15; Length 742;
Best Local Similarity 21.2%; Pred. No. 2.6;
Matches 66; Conservative 49; Mismatches 134; Indels 63; Gaps 10;

QY 18 ETAGALDFYKYLDOVIPKTFDTEIKELSRFKQEQYQASVLVGDIKVLLMDSQDKYF 77
DB 236 KTVDLASDATSKLSDTVKDKISDLPTIKT-----LNDTKLLSSDLKFLDFTNDLD 288
QY 78 EATQTVVEWCQVVTQLLSAYILLDFEYNKKASQAQDKILIRLD-----DGVNKLNEAQ 131
DB 289 ELSPLIKSDMLMDVL-----SSSASSUTLMLIDAVNSGSEVDPKLDNL 333
QY 132 KSLIGSSQSFNNSGKLLALDSQTN-----DFSEKSSYFQSQVDRIKREAYA 179
DB 334 SEKLSNLQSLNDTLVDFTLKNQLTSNNRLDDVDNLEDSSNKIDSSITLNDIKNKVIS 393
QY 180 GAAGIVAGPGLIISYIAAGVIEGKLIPELNDRL-KAVQNFTSLSVTKQANKDI-- 236
DB 394 GQOPSISA--LNNVLSNGIGRINLNNFDSKISKPINNFAN---SIKVANDIITV 448
QY 237 -DAAKLK-----LATEIAAIGETKETEITFRFYVDYDDLMLSLKGAAGKAMINTCNEY 288
DB 449 LDKAEAKLPKVEILTTSLKSGNAQSISLIR-----ERLPLAKGMLDLDLITLSKI 501

SEQ ID NO 5109
LENGTH: 1938
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-369-493-5109

Query Match 7.1%; Score 106.5; DB 14; Length 1938;
Best Local Similarity 17.3%; Pred. No. 12;
Matches 52; Conservative 63; Mismatches 112; Indels 73; Gaps 8;

QY 5 FAEQTVVVK-----SAIETADGALDFYKYLDOVIPWKTFTDETIKELSRFK 51
DB 1315 FSSQLVEAKKAAAEDELHERQEFHAAACKNLEHLEDOCHELLEEQINGK--DDIQQLSRIN 1372
QY 52 QEYSQ-----EASVLVG--DIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEY 104
DB 1373 SEISQWKARYEGEGLVGESELEELKQKQNRVMDI-QEALSAAQNKVISLEKAKGLLAET 1432
QY 105 NEKKASAQKDILI-----RIIDDGVNKLNEAQSLLGSSQSFNNASGKLALD 152
DB 1433 EDARSVDRLHVTIASLEKKQRAFDKIVDDMKRVDDIQKEIDATTRDSRNTSTEVFKLR 1492
QY 153 SOLTNDSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGILIIISYIAAGVIEGKLIPELN 212
DB 1493 SSMDN-----LSEQIETLRRE-----NKIFSQBIRDIN 1520
QY 213 DRLKAVQNFTSLSVTVKQAKND-----IDAAKKLATEIAAIGEIKTETETTRFYVD 265
DB 1521 EQITQGGRTYQEVHKSVRRLQEKDELOHALDEAAEALEAESKVLRLQIEVQOIRSETE 1580

RESULT 21
US-10-369-493-5110
Sequence 5110, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 5110
LENGTH: 1938
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-369-493-5110

Query Match 7.1%; Score 106.5; DB 14; Length 1938;
Best Local Similarity 17.3%; Pred. No. 12;
Matches 52; Conservative 63; Mismatches 112; Indels 73; Gaps 8;

QY 5 FAEQTVVVK-----SAIETADGALDFYKYLDOVIPWKTFTDETIKELSRFK 51
DB 1315 FSSQLVEAKKAAAEDELHERQEFHAAACKNLEHLEDOCHELLEEQINGK--DDIQQLSRIN 1372
QY 52 QEYSQ-----EASVLVG--DIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEY 104
DB 1373 SEISQWKARYEGEGLVGESELEELKQKQNRVMDI-QEALSAAQNKVISLEKAKGLLAET 1432
QY 105 NEKKASAQKDILI-----RIIDDGVNKLNEAQSLLGSSQSFNNASGKLALD 152
DB 1433 EDARSVDRLHVTIASLEKKQRAFDKIVDDMKRVDDIQKEIDATTRDSRNTSTEVFKLR 1492
QY 153 SOLTNDSEKSSVFQSDVRIRKEAYAGAAAGIVAGPFGILIIISYIAAGVIEGKLIPELN 212

289 QQRHGKKTLLV 300
502 SNGEDMKLVSL 513

RESULT 19
US-10-369-493-9374
Sequence 9374, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 9374
LENGTH: 857
TYPE: PRT
ORGANISM: Xylella fastidiosa
US-10-369-493-9374

Query Match 7.1%; Score 107.5; DB 14; Length 857;
Best Local Similarity 23.9%; Pred. No. 3.2; Indels 104; Gaps 14;
Matches 68; Conservative 45; Mismatches 104; Indels 67; Gaps 14;

QY 7 EQTVVVKSAIETADGALDFYK-----YLDQVIPWKTFTDETIKELSRFKQEYQASV 60
DB 574 EBAIKVSDAVRSRTGLSDPNRPSGFLFLGTGVGKT--ELCKALAEFLPD-SQDAMV 630
QY 61 LVGDIKVLLMDSDQKYFEAT-----QTVYEWCGVVTQLLSA---YILLFDEYNEKKAS 110
DB 631 -----RIDMSEFMEKHSVARLIGAPPYGYEGGYTELVRPYPYSLILLDEV-EKAS 684
QY 111 AQKDILRIIDDGVNKLNEAQSLLGSSQSFNNASGKLALDLSQLTNDSEKSSVFQSQV 170
DB 685 DVNILLQVLDG--RLTDCQ-----GRTVDFRNT---VIVMTSLGSHQIQELSGDDS-- 733
QY 171 DRIRKEAYAGAAAGIVAGPFGILIIISYIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVK 230
DB 734 ---PEVVTQKAAVM-----GVVQAHPRPFINRLDDIVVF----- 766
QY 231 QANKDIDAANKLATEIAAIGEIKTETETTRFYVDYDMLSL 274
DB 767 ---HPLDKAQIKQIARIQLRGLERLAE--SELKLDLDDRALELL 806

RESULT 20
US-10-369-493-5109
Sequence 5109, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
NUMBER OF SEQ ID NOS: 47374

Db 1493 SSMON-----LSEQIETURR-----NKIPSOEIRDIRN 1520

Qy 213 DRLKAVQNFYSLSVTKVQANKD-----IDAAKLKLATEIAAIGETETTRFYVD 265

Db 1521 EQITQGGRTYQEVHKSVRRLQKDEQALDEAAEALESKVLRRLQIEVQIRSEIE 1580

RESULT 22

US-10-282-122A-71235

; Sequence 71235, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 71235

; LENGTH: 3692

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-282-122A-71235

Query Match 7.1%; Score 106.5; DB 15; Length 3692;

Best Local Similarity 21.5%; Pred. No. 28;

Matches 71; Conservative 49; Mismatches 129; Indels 81; Gaps 12;

Qy 11 EVKSAIETADGALDFYKYLQVVPKTFDTIKELSR-----FKQEYSQ-EA 58

Db 564 EQVKNDDIIPSNYTLASYNKY-----NKLKRAQTVLDEETNTPFNQRYSTQI 612

Qy 59 SVLVGDIKVLMM-----DSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASA 111

Db 613 DDLHLLEQTLLINRVRSASREINDKAQEMTDAYDSTELTTEKDT---LVDQIENHNEI 669

Qy 112 QKDILIRILDGCVKLNKAQKSLGSS-----QSFNNAAGKLLALDSQLTNDIFS 160

Db 670 SNNIDDELTDGCVKKEAGLHLESDTFPHVTKFNARQVNNRA-----DQCKTLIRN 723

Qy 161 EKSYFQSOVDRIK-EAYAGAAAGIVNGPGLIISIAAGVIEGKLIPELNDELKAVQ 219

Db 724 NHEATTBEQNEAIRQVEAHSSDA-----IAKIGEAETDTTWN-----ARD 764

Qy 220 NFFTSLSVTKVQANKDIDAAKLKLATEIAAIGETETTRFYVDYDDLMLSLKGAAK 279

Db 765 NGTKLIATDVNPNTKKAEEA---RAAVTNSANSKIDNNNTQATLDERDAIALVNRSD 821

Qy 280 KMINTCNEYQ-----QRHGKKTLLVVP 301

Db 822 EAIQNINTAQGNDDVTEAQNNGTNTTIOQVP 851

RESULT 23

US-10-028-248A-105

; Sequence 105, Application US/10028248A

; Publication No. US20030235882A1

; GENERAL INFORMATION:

; APPLICANT: Shinkete, Richard

; APPLICANT: Patturajan, Meera

; APPLICANT: Vernet, Corine

; APPLICANT: Casman, Stacie

; APPLICANT: Malyankar, Uriel

; APPLICANT: Shenoy, Suresh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gangolli, Esha

; APPLICANT: Miller, Charles

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Smithson, Glennda

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Liu, Xiaohong

; APPLICANT: Colman, Steven

; APPLICANT: Tchernev, Velizar

; APPLICANT: Si, Jingsheng

; APPLICANT: Edinger, Shlomit

; APPLICANT: Stone, David

; APPLICANT: Sciore, Paul

; APPLICANT: Millet, Isabelle

; APPLICANT: Rothenberg, Mark

; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods

; FILE REFERENCE: 21402-222

; CURRENT APPLICATION NUMBER: US/10/028,248A

; CURRENT FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/256619

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/262959

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/272408

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/285189

; PRIOR FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: 60/308039

; PRIOR FILING DATE: 2001-07-26

; PRIOR APPLICATION NUMBER: 60/311266

; PRIOR FILING DATE: 2001-08-09

; NUMBER OF SEQ ID NOS: 211

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 105

; LENGTH: 1961

; TYPE: PRT

; ORGANISM: Rattus norvegicus

US-10-028-248A-105

Query Match 7.0%; Score 106; DB 14; Length 1961;

Best Local Similarity 21.7%; Pred. No. 13;

Matches 68; Conservative 50; Mismatches 100; Indels 96; Gaps 15;

Qy 7 EQTVEVKSATETADGALDFYKYLQVVPKTFDTIKELSRFKQEYSQEAASVLGVGDIK 66

```
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 105
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-105

Query Match          7.0%; Score 106; DB 15; Length 1961;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 68; Conservative 50; Mismatches 100; Indels 96; Gaps 15;

Db      1166 EQEVSILKKTLE-----DEA---KTHEAQIQEM---ROKHSQAVBELAEQLE 1206
QY      67 VLLMDSQKYFEAT-----QTVYEWCGVVTQLLSAYILL-----FDEYNEKKAQAQ-KDIL 116
Db      1207 -----QTKRVKATLEKAKOTLENERGELANEVKA---LLOKGDSEHKKRKKVEAQLOELQ 1258
QY      117 IRI-----LDDGVNKLNEAQKSLGSSQSFNNASGKLL-----ALDSQLTN----- 157
Db      1259 VKFSEGERVTELADKVKSLQVELDSVTGLNQSDSKSLTKDPSALESQLODTQELLQ 1318
QY      158 -----DFSEKSSYFQSQVDRIKREAYAGAAAGIVAGPFGLLIISYIAAGVI 203
Db      1319 EENRQKLSLSTKLQWEDKNSFRQLEEBEERKNEKQIAT----- 1362
QY      204 EGKLIPELNDRLKAVONFTSLSVTVKQA---NKDIDAAKKLATEIAAIGEIKTETET 259
Db      1363 ---LHAQVTDMMKKMEDGVGCLE-TAEAEKRRLOKDLGLSQRLEEKVAAYD--KLETK 1416
QY      260 TRFYVDYDDLMLSL 273
Db      1417 TRLQQLDLDLVL 1430
```

```
RESULT 24
US-10-107-782-105
; Sequence 105, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Steve,
; APPLICANT: Colman, Steve,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Gangolli, Bsha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li, Xiaohong,
; APPLICANT: Liu, Xiaohong,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zethusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT FILING DATE: 2002-03-27
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2001-01-19
; PRIOR FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2001-04-20
; PRIOR FILING DATE: 2001-07-26
; PRIOR FILING DATE: 2001-07-26
; PRIOR FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
```

```
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 105
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-105

Query Match          7.0%; Score 106; DB 15; Length 1961;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 68; Conservative 50; Mismatches 100; Indels 96; Gaps 15;

Db      1166 EQEVSILKKTLE-----DEA---KTHEAQIQEM---ROKHSQAVBELAEQLE 1206
QY      7 EOTVEVKSATETADGALDFYKYLQVTPKWTDFETIKELSRFKQESQASVLVGDIK 66
Db      1166 EQEVSILKKTLE-----DEA---KTHEAQIQEM---ROKHSQAVBELAEQLE 1206
QY      67 VLLMDSQKYFEAT-----QTVYEWCGVVTQLLSAYILL-----FDEYNEKKAQAQ-KDIL 116
Db      1207 -----QTKRVKATLEKAKOTLENERGELANEVKA---LLOKGDSEHKKRKKVEAQLOELQ 1258
QY      117 IRI-----LDDGVNKLNEAQKSLGSSQSFNNASGKLL-----ALDSQLTN----- 157
Db      1259 VKFSEGERVTELADKVKSLQVELDSVTGLNQSDSKSLTKDPSALESQLODTQELLQ 1318
QY      158 -----DFSEKSSYFQSQVDRIKREAYAGAAAGIVAGPFGLLIISYIAAGVI 203
Db      1319 EENRQKLSLSTKLQWEDKNSFRQLEEBEERKNEKQIAT----- 1362
QY      204 EGKLIPELNDRLKAVONFTSLSVTVKQA---NKDIDAAKKLATEIAAIGEIKTETET 259
Db      1363 ---LHAQVTDMMKKMEDGVGCLE-TAEAEKRRLOKDLGLSQRLEEKVAAYD--KLETK 1416
QY      260 TRFYVDYDDLMLSL 273
Db      1417 TRLQQLDLDLVL 1430

RESULT 25
US-10-369-493-497
; Sequence 497, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 473/74
; SEQ ID NO 497
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-497

Query Match          7.0%; Score 105.5; DB 14; Length 344;
Best Local Similarity 21.2%; Pred. No. 1.4;
Matches 56; Conservative 53; Mismatches 102; Indels 53; Gaps 13;

Db      15 SALETADGALDFYKYLQVTPKWT--FDETIKELSRFKQESQASVLVGDIKVLMDS 72
QY      71 TVLHVGDSSIRLFHQPDISDLFWKELGIDIVLDCTGKYGERIDGSEAHLASGAKKVLFAHP 130
Db      73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDI---LIRILDDGVNKLNE 129
QY      131 GGNLDLAT-VVY---GVNHELLSA-----EDCIVNASCTTNCIIPILKVLDD----- 174
Db      130 AQKSLGSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDRIKREAYAGAAAGIVAGP 189
```

Db 175 -----AFNIESGVTTHASHM-QNDQVIDAY-HSDLRTR-----AASQSIIP-- 215
Qy 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVONFSTLSVTVKQAKDIDAOKLKLATEIAA 249
Db 216 ----VDTKLAAGIT--RIFPKFSNFEALSVRVPINVTAILDSVTVSA-----AVNVSE 264
Qy 250 IGEI--KTETETTRFYVDVDDLML 271
Db 265 VNQLLKSAVSSFRGIVDFDTLPL 288

RESULT 26

US-10-437-963-107270
; Sequence 107270, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 107270
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11638C.1.pap
US-10-437-963-107270

Query Match 7.0%; Score 105; DB 16; Length 737;
Best Local Similarity 21.3%; Pred. No. 4.3;
Matches 69; Conservative 63; Mismatches 126; Indels 66; Gaps 12;

Qy 4 IPAEQTVVVKSAITAD---GALDFYKYLQVPIWKTFD---ETIKLSRFRKQSYQ 56
Db 249 LVAEQKLIACEAIEERKMLGALTEANE-----AAAKAFDTQNEBITKELEDLTKLEE 303
Qy 57 -----EASVLGDIKVLMDSQDYFEATQTVYVWCGVVTQLLSAYILLFDEYNEKAS 110
Db 304 IKTNKDLAESENGKLSRSELLSABEYKSYSEAEV---KYLKQVMGA-VVEAKEAAAKAFA 358
Qy 111 AQKDILIRLDDGVNKLNEAQSLLGSSOSFNNAAGKLIALDSQLTNDPSEKSSYFQSQV 170
Db 359 AEKEDIKESDNLKRVKEIQDLSKLLVSENDLASEILSM-----KQKHGQFEVEV 410
Qy 171 DRIRKEAYAGAAGIVAGFGLIISYSIAAGVIEGKLIPELNDRLKAVONFSTS----- 224
Db 411 TSLKKELGALEAKEIT-----TRAFEVKEITLKELEDLKRKVVIEIQTNKDLVEV 461
Qy 225 -----LSVTVKQAKDIDAOKLKLATEIAAIGEIKETETTRFYVDVDDLMLSLK 275
Db 462 GNDKLRDLVLSAQKQKQSTLEEVANNLKM--ELGALVEAK-EVATKAFDAEKAKI----- 512
Qy 276 GAAKMINTCNEYQQRHGKKTLE 299
Db 513 --TKELEDVKRWKEIQVKDLVE 534

RESULT 27

US-09-815-242-10796
; Sequence 10796, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10796
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10796

Query Match 6.9%; Score 104; DB 9; Length 722;
Best Local Similarity 19.8%; Pred. No. 5;
Matches 66; Conservative 51; Mismatches 109; Indels 108; Gaps 12;

Qy 16 AIEVADGALDFYKYLQVPIWKTFDETIKLSRFRKQSY-----SQE 57
Db 122 AVKGNDKIKSYDDLLKGTVAAKVGTESANFLEKNKEKYDYTIKNFDDATGTYKALENGE 181
Qy 58 ASVLGDIKVL-----LMDSQD-----KYFEATQTVYVWCG 88
Db 182 ADAIVDDYPVLGYAVKNGQKLQVGDKETGSSYFAVKKQNPelikfknaglnkng 241
Qy 89 VVTQLLSAYILLFDEYNEKASAKDILIRLDDGVNKLNEAQSLLGSSQS-----PN 142
Db 242 TYDKILNNYLATGDETNTQDAGEQ-----MKKITPKKEYVIASDSTFAPFEFQ 290
Qy 143 NASGKLALDQLTNDPSEKSSY-----FQSOVDRIKEAYAGAAGIVAGFPGLII 194
Db 291 NAQGDYVIDVLLVKRAAELOQTFVEFKFIFGSSAVQAVE-----SGQADGWAG----- 340
Qy 195 SYSTAAGVIEGKLIPELNDRLKA-----VONFSTLSVTVKQAK-----KQIDAAK-----LK 242
Db 341 -----NTITDRKKAFDFSVPYFDSGFIQIAVKKGNDKIKSYDIDLKGGKVGK 387
Qy 243 LATEIA-AIGIKETETTRFYVDVDDLMLSLK 275
Db 388 IGTSADFLKNKKKYDYSIKYLTDTTDLALYSALE 421

RESULT 28

US-09-841-260-139
; Sequence 139, Application US/09841260
; Publication No. US20030175700A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; APPLICANT: Stromberg, Erika Jean
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS

us-09-993-292b-24.rapb

Wed Feb 2 11:26:41 2005

Db 268 LQAQNNSPDNIAATKELIDAAETKVNELKQEHGTL-----TDSPLVKK 311

QY 159 FSEKSSYFQSOVDRIKEAVAGAAAGIVAGPGLIISYIAAGVIEGK-----LIPE 210

Db 312 AEEQISOAQKDIOEIKP---SGSDIPV-GPSGSAAGSAGALKSSNNSGRISLLDD 367

QY 211 LNDRLKAV-----QNFFTSLSVTVKQANKDIDAAKLKATE-----IAAIGEIKTETETTR 261

Db 368 VDNEMAIALQGFMSIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGDGELPAEIOAIK 427

QY 262 FYVDYDDLMLSLK 275

Db 428 -----DALAQALK 435

RESULT 30

US-10-762-058-139

; Sequence 139, Application US/10762058

; Publication No. US20040137007A1

; GENERAL INFORMATION:

; APPLICANT: Bhatia, Ajay

; APPLICANT: Probst, Peter

; APPLICANT: Stromberg, Erika Jean

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS

; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.515D1

; CURRENT APPLICATION NUMBER: US/10/762,058

; CURRENT FILING DATE: 2004-01-15

; NUMBER OF SEQ ID NOS: 140

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-10-762-058-139

Query Match 6.9%; Score 103.5; DB 16; Length 660;

Best Local Similarity 21.7%; Pred. No. 4.9;

Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTVEVVKSAIETADGALDFYNYL-----DQV-----IPWKTFF 40

Db 155 EVNNIKKALEAQKDTIDKLNKLVTLQNKSKLTVLKTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQYSEASVLDGDIKVLMDSDQ---KYFEATQTVYEWGVVTVQLLSAYI 98

Db 215 DQIHKDLER--QNISYEA-VLTNAGEVIKASSEAGIKLQALQSIVD-AGDQSQ---AAV 267

QY 99 LLFDEYNEKKASAKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTND 158

Db 268 LQAQNNSPDNIAATKELIDAAETKVNELKQEHGTL-----TDSPLVKK 311

QY 159 FSEKSSYFQSOVDRIKEAVAGAAAGIVAGPGLIISYIAAGVIEGK-----LIPE 210

Db 312 AEEQISOAQKDIOEIKP---SGSDIPV-GPSGSAAGSAGALKSSNNSGRISLLDD 367

QY 211 LNDRLKAV-----QNFFTSLSVTVKQANKDIDAAKLKATE-----IAAIGEIKTETETTR 261

Db 368 VDNEMAIALQGFMSIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGDGELPAEIOAIK 427

QY 262 FYVDYDDLMLSLK 275

Db 428 -----DALAQALK 435

RESULT 31

US-10-282-122A-58016

; Sequence 58016, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.515

; CURRENT APPLICATION NUMBER: US/09/841,260

; CURRENT FILING DATE: 2001-04-23

; NUMBER OF SEQ ID NOS: 140

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-841-260-139

Query Match 6.9%; Score 103.5; DB 10; Length 660;

Best Local Similarity 21.7%; Pred. No. 4.9;

Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTVEVVKSAIETADGALDFYNYL-----DQV-----IPWKTFF 40

Db 155 EVNNIKKALEAQKDTIDKLNKLVTLQNKSKLTVLKTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQYSEASVLDGDIKVLMDSDQ---KYFEATQTVYEWGVVTVQLLSAYI 98

Db 215 DQIHKDLER--QNISYEA-VLTNAGEVIKASSEAGIKLQALQSIVD-AGDQSQ---AAV 267

QY 99 LLFDEYNEKKASAKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTND 158

Db 268 LQAQNNSPDNIAATKELIDAAETKVNELKQEHGTL-----TDSPLVKK 311

QY 159 FSEKSSYFQSOVDRIKEAVAGAAAGIVAGPGLIISYIAAGVIEGK-----LIPE 210

Db 312 AEEQISOAQKDIOEIKP---SGSDIPV-GPSGSAAGSAGALKSSNNSGRISLLDD 367

QY 211 LNDRLKAV-----QNFFTSLSVTVKQANKDIDAAKLKATE-----IAAIGEIKTETETTR 261

Db 368 VDNEMAIALQGFMSIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGDGELPAEIOAIK 427

QY 262 FYVDYDDLMLSLK 275

Db 428 -----DALAQALK 435

RESULT 29

US-10-007-693-139

; Sequence 139, Application US/10007693

; Publication No. US20020146776A1

; GENERAL INFORMATION:

; APPLICANT: Bhatia, Ajay

; APPLICANT: Probst, Peter

; APPLICANT: Stromberg, Erika Jean

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT

; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.515C2

; CURRENT APPLICATION NUMBER: US/10/007,693

; CURRENT FILING DATE: 2001-12-05

; NUMBER OF SEQ ID NOS: 157

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-10-007-693-139

Query Match 6.9%; Score 103.5; DB 13; Length 660;

Best Local Similarity 21.7%; Pred. No. 4.9;

Matches 68; Conservative 48; Mismatches 119; Indels 79; Gaps 14;

QY 8 QTVEVVKSAIETADGALDFYNYL-----DQV-----IPWKTFF 40

Db 155 EVNNIKKALEAQKDTIDKLNKLVTLQNKSKLTVLKTDSADQIPAINSOLEINKNSA 214

QY 41 DETIKELSRFKQYSEASVLDGDIKVLMDSDQ---KYFEATQTVYEWGVVTVQLLSAYI 98

Db 215 DQIHKDLER--QNISYEA-VLTNAGEVIKASSEAGIKLQALQSIVD-AGDQSQ---AAV 267

QY 99 LLFDEYNEKKASAKDILIRILDDGVNKLNEAKSLGSSQSFNNASGKLLALDSQLTND 158

APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 58016

LENGTH: 1020

TYPE: PRT

ORGANISM: Enterococcus faecium

US-10-282-122A-58016

Query Match 6.9%; Score 103.5; DB 15; Length 1020;

Best Local Similarity 19.8%; Pred. No. 8.9;

Matches 65; Conservative 41; Mismatches 90; Indels 133; Gaps 12;

Qy 3 GIFAQQTVEVVK-----SATETADG---ALDFYKYLDOVIPWKTPTDETIELSR 49

Db 159 GIF-EAAGVLKYKQKKAQKLPETEDNLSRVQDIIHELEEQUTPLAAQSEAAKEFLR 217

Qy 50 FQEQYSQ-EASVLVGDIVKLLMDSQDKYFEATQTVYWGCVVTVQLLSAYILLFDYNEKK 108

Db 218 LKETLTQTDVSLMVAEIKTKKD-----WDNQ 245

Qy 109 ASAQKDIILRILDDGVNKLNER---QKSLGSSQSFNNAAGKLLALDSQLTNDPSEKSSY 165

Db 246 AQLAKFNL-----ELGKLSIESIQQESILAKQRKENAQADRLIEKNOQVLLDLSEKLQ 299

Qy 166 FQSOVDRIKRAYAGAAAGVAGPGGLIISYIAGVIEGKLIPELNDRLKAVQNFFTSL 225

Db 300 TEGQKD-----VLQERTKHTQKSGQY 321

Qy 226 SVTVQANKNDIAAKLKLATEIAAIGEIKTETETTRFYVDYDDMLSLKGAARK----- 280

Db 322 QTSLAEAK-----KVK-----HFEKLQESLMKAAAEKETEIQ 354

Qy 281 -----MINTCNEYQ--QRHGKTKLLEVPD 302

Db 355 KAEANLIKTQBLEKYQKSTKELLAELRD 383

RESULT 32

US-10-282-122A-51864

; Sequence 51864, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 51864

LENGTH: 1163

TYPE: PRT

ORGANISM: Clostridium acetobutylicum

US-10-282-122A-51864

Query Match 6.9%; Score 103.5; DB 15; Length 1163;

Best Local Similarity 21.7%; Pred. No. 11;

Matches 81; Conservative 55; Mismatches 142; Indels 95; Gaps 14;

Qy 1 MTGIFAEQTVVVKSAITETADG---ALDFYKYLDOVIPWKTPTDETIK-----ELS 48

Db 112 ITGDEVEVLGSAKSVNEKQOEIIGLSLDDTRTV--VLPOCKFSEFLKLEKERNMLE 169

Qy 49 RF--KQEYSQSEAS-----VLVGDIVKLLMDSQDKYFEATQTVYWGCVVT 91

Db 170 RLFLNQLQYGDLSFKLARKIRKEREKENVLGVNEDVLKERELLKENDDFN 229

Qy 92 QLLSAYILLFPEYNEKKAQAKDILI-----RILDDGVNKLNE---AQKSLIGSSQS--- 140

Db 230 EASKEYLKABEYNEKGVWGLQIEIEKNRVKDLMEKDEIDLKEKRALGESSKVK 289

Qy 141 -----FNNASGKLALDSQLTN-----DFSEKSSYFQSDVRIRKAYAGAA 182

Db 290 PYINDYENTLKQIDILKEQILSRENTWKAI SLEKDEKMKLSIAKDNKEK----- 339

Qy 183 AGIVAGPFGLLIISYIAGVIEGKLIPELNDRLKAVQNFFTSLSV-----TVQAKND 235

Db 340 ----ALPKFMKHHIILDAIKEKDLDDNIKLEKKRLQKIEKLSLEASKEELIKQIKD 395

Qy 236 IDAAKCLKLATEIAAIGEIKTETETTRFYVDYDDMLSLKGAAKKNIN-----TCNEYQ 289

396 IDSLTLKIQNLESKIDNLKVPE-----YKKNKINEGIFLLRNYDEKLKHKNKGLGDCDKFKQ 451
Db
396 IDSLTLKIQNLESKIDNLKVPE-----YKKNKINEGIFLLRNYDEKLKHKNKGLGDCDKFKQ 451
QY
290 ----QRHGKTKLL 298
Db
452 VDFEAKSKKEMWL 464
Db
RESULT 33
US-10-408-765A-741
; Sequence 741, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 741
; LENGTH: 5373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-741
Query Match 6.9%; Score 103.5; DB 16; Length 5373;
Best Local Similarity 23.0%; Pred. No. 86;
Matches 56; Conservative 49; Mismatches 106; Indels 33; Gaps 10;
QY 29 KYLDQVTPKTFDTETIKELSRKQYSQASVLVGDIKVLLMSDQDKYFEATQTVYWCW 88
Db 1369 KYISDAL--RLREEKVEEKEQHEVKVKGWVSTLARNTQK---ATSETKEST 1423
QY 89 VVTQLLSAYILLFDEVNEKK---ASAKOILIRILDDGVNKLNEAKSLLGSSQSFN--- 142
Db 1424 DIEKAILEQQVLSEELTTKEQVSEAIKTSQIFLAKHG-HKLESEKQKI--SEQNALN 1480
QY 143 -----NASGKLALDSOLTNDPSEKSSYFQSDVRIRKE-----AYAGAAAGIVAGP 189
Db 1481 KAYHDLCDGSANQQLQSQSLAHQTEQKT--LQKQNTCHQOLEDLCSWVGQAERALAGH 1538
QY 190 FGLIISYSIAAGVIEGKLIPELNDRLKAVQNFTSLSVTVKQANKDIDAKLKLKLA-TEIA 248
Db 1539 QGRITQODLSA---LQKNQSDLKDLQDDIQNRATSFATVVKVDIEGFMEENQTKLSPRELT 1595
QY 249 AIGE 252
Db 1596 ALRE 1599
RESULT 34
US-10-369-493-19762
; Sequence 19762, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19762
; LENGTH: 862
; TYPE: PRT
; ORGANISM: NO. US20030233675A1toc punctiforme
US-10-369-493-19762
Query Match 6.8%; Score 103; DB 14; Length 862;
Best Local Similarity 23.7%; Pred. No. 7.8;
Matches 79; Conservative 41; Mismatches 111; Indels 102; Gaps 17;
QY 38 KTFDETIKELSRKQYSQASVLVGDIKVLLMSDQDKYFEATQTVYWCWGVVTTOLLSAY 97
Db 61 KLDRTEQFFQRPQKVSQGTSSYILGRSLDTLLDRADYVRKEFQDEYL---SIEHLLAY 117
QY 98 I-----LLFDEVNEKKASAKOILIRILDDGVNKLNEAKSLLGSSQ-SFNNASGKLL 149
Db 118 AKDRFGKALFOFG-----LDSE--KLKDIKQVRGSKVTDQNPGEKYE 161
QY 150 ALDSQLTNDPSEKSSYFQ-----SQVDRIR-----KEAYA-G 180
Db 162 ALB-KYGRDLTEAARGQLDPVIGRDEIRRTVQILSRRTKNNPVLIGEPGVGKTAIAEG 220
QY 181 AAAGIVAG--PFGL-----IISYSIAAGVIEGKLIPELNDRLKAVQNFT---SLSVTVK 230
Db 221 LAQRITAGDVPOSLKDKRGKISLDMGALIAKAPRGEFEELKAVLKEVTESGNNVLFID 280
QY 231 QANKDIDAALK-----LATEIAAIGIKTETTT-----RF---YV 264
Db 281 EIHVVAGATQGMADAGNLLKPLMARGELRCIGATTLDBYRKHIEKDAALERRFOQVYV 340
QY 265 DYDDL--MLSLKGAAKKMMINTCNEYQORHGKK 295
Db 341 DQPSVEDSISILGLRER-----YENHGVK 366
RESULT 35
US-10-080-608A-66
; Sequence 66, Application US/10080608A
; Publication No. US20030199956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-080-608A-66
Query Match 6.8%; Score 103; DB 14; Length 961;
Best Local Similarity 22.0%; Pred. No. 9.1;
Matches 70; Conservative 53; Mismatches 129; Indels 66; Gaps 14;
QY 11 EVVKSIAETADGALDFYNNYL-----DQVTPKTFDE-----TIKELSRFK 51
Db 639 EEVKKTLQEHDSIVTHYKNMIREQDLLEELKQKQISTLKQCNQEQLTAVTQVQSQIQOHK 698
QY 52 QEYSQASVLVGDIKVLL--MSODK--YFEATQTVYWCWGVVTTOLLSAYILLFDEVNEKK 108
Db 699 DOYNL-----LKVQLGKOSQHQGPYTDGAG-----MNGVQPEISR---LREEIEELK 743
QY 109 AS-----ACKDILIRILDDGVNKK--LNEAQSLLGSSQSFNNASGKLALDSQLND 158
Db 744 SNRELQSLAEKSLIENLKSSQLSPGTNEQSATAGDSEIQIAELKQELATLKSQ-NS 802

QY	159	FSEKSYFOSQVDRTRKAYAGAAAGIVAGPGLIIISYSTAAGVIECKL-----IPELN 212
DB	803	QSVETIKLQTEKQELQKLTQTEAFKSAAPVGESETVIATKTD--VEGRSALLQETKEUK 860
QY	213	DRLKAVQNFFTSLSVTVQKANKDIAAKLKLATSIAGAIEKTTETETRTFRVYDDMLMS 272
DB	861	NEIKALSERTAIKQLDSSNSTI-----AILQNEKNKLEVDITDSKKEQDDLLV- 910
QY	273	LLKGAAKWMTNCNEYQQ 290
DB	911	LLADODOKIFSLKMKLKE 928

RESULT 36
 US-10-370-685-155
 ; Sequence 155, Application US/10370685
 ; Publication No. US20030215903A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyman, Paul
 ; APPLICANT: Goldberg, Edward
 ; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
 ; FILE REFERENCE: NANF_P-004
 ; CURRENT APPLICATION NUMBER: US/10/370,685
 ; CURRENT FILING DATE: 2003-02-21
 ; PRIOR APPLICATION NUMBER: 10/080,608
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 159
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 155
 ; LENGTH: 961
 ; TYPE: PRT
 ; ORGANISM: Bos taurus
 US-10-370-685-155

Query Match	6.8%;	Score 103;	DB 14;	Length 961;	
Best Local Similarity	22.0%;	Pred. No. 9.1;			
Matches	70;	Conservative 53;	Mismatches 129;	Indels 66;	Gaps 14;
QY	11	EVVKSALETADGALDFYKNYL	-----DOVIPWKTFDE-----TIKELSRFK	51	
DB	639	EEVKTLIEQHDSIVTHYKNIREQDLQLEELKQISILKQNEQLQTAVTQVQSQIQHKK	698		
QY	52	QEYSQEASVLVGDIKVLL-MDSQDK--YFEATQTVYEMCGVVTQLLSAYILLDFEYNEKK	108		
DB	699	DQYNL-----LKVQLGKDSHQGFYDGAQ-----MNGVQPEEISR---LREEIELK	743		
QY	109	AS-----AQKDILRIILDDGVNK--LNEAQSLIGSSQSFNNAGSKLLALQSOLTND	158		
DB	744	SNRELLQSALAEKDSLIENLKSSQLSPGTNEQSATAGDSQIAELKQELATLKSQ--NS	802		
QY	159	FSEKSSVFQSQVDRIKREAVAGAAGIVAGPFGILLISVSIAGVIEGKL-----IPELN	212		
DB	803	QSVETITKLQTEKQELLQKTAFAKSAVPGBSESVIATKTTD--VEGRLSALLOETKELK	860		
QY	213	DRUKVQNFVTSLSVTVKQANKDIDAARKLATEIAAIGEIKTETETTRFFVVDVDDMLMS	272		
DB	861	NEIKALSEERTAIKEQLDSSNSTI-----AILQNEKNKLEVDITDSKEQDDLLV-	910		
QY	273	LLKGAACKMINTCNEYQQ	290		
DB	911	LLADODOKIFSLKNKLKE	928		

RESULT 37
US-10-408-765A-1635
; Sequence 1635, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boi D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.

```

; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1635
; LENGTH: 1583
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1635

```

Query Match	6.8%;	Score 103;	DB 16;	Length 1583;
Best Local Similarity	19.9%;	Pred. No. 18;		
Matches	73;	Conservative 55;	Mismatches 89;	Indels 150;
Gaps	16			
Qy	7	EQTVEVVKSAIETADGALDFYKNYLDQVWPKTT--FDETIKELSRFKOEYSQEASVLVGD	64	
Db	531	EQTIQ-----YNSELEQKVNELTGGLEETLKE-----KDQNDQKLEKLMVQ	571	
Qy	65	IKVLLMDSODKYFEATQTVYWCWGVQLLSAYI--LLFDEYNEKKAQAKQILIRILD--	121	
Db	572	MKVL-----SEDK-----EVLSAEVKSLYEENN--KLSEKKQLSRDLEVF	610	
Qy	122	-----DGVNKL--NEAQKSLLGSSQSFNNASGKLLALD	152	
Db	611	LSOKEDVILKEHITQLEKQLQLMVEQDNLNKLLENEQVKQLFKVTQLY-----GFLKEMG	666	
Qy	153	SQLTNDFSEKSSYFQSQVDRIKREAYAGAAAGIVAGPGLLIISYIAAGVIEGKL-----	207	
Db	667	SEVSEDSSEKD-----VVNVLQAVGESLAKINESEKC	697	
Qy	208	-----IPELNDRLKAVQNPFSTLSVTVVQANKOIDAAKLKLATIPAAIG-----	251	
Db	698	NLAFQDEKVLKEKEIKCLQEESVWQCEELKSLLRDYEQEKVLLRKELEETQSEKEALQ	757	
Qy	252	-----EIKTETETFRFYVDLMLSLKGAAKMINTCNEYQQRHGK-----KT	296	
Db	758	SDLLEKMANEKTR--LENQNLIIQV-----EVSQTCSEIHNKEKCFIKHEHNLP	810	
Qy	297	LLEVVDI	303	
Db	811	LASOKET	817	

```

RESULT 38
US-10-282-122A-55283
; Sequence 55283, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 961c-741
; US-10-220-480-50

Query Match      6.8%; Score 102; DB 15; Length 590;
Best Local Similarity 21.4%; Pred. No. 5.7;
Matches 74; Conservative 62; Mismatches 129; Indels 80; Gaps 17;

QY 5 FAEQTV-----EVVKSIAETADGALDFYKYLDOVIPWKT-FDETIKELSRFKQEQSQ 56
Db 131 FAEETKTIVIKIDEKLEAVADTVDKHAEAFNDIADSLDETNTKAEAVKTANEAKQT-AE 189
QY 57 EASVLGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYI-----LLFDEYNEKKAAS 110
Db 190 ETKQNV-DAKVKAETAAGKAEAAAGTANTAADKAEAAKVTDIKADIATNKDNIKA 248
QY 111 AQKDIILIR-----ILDDGVNKLNEAOKSLGSSQ-SFNNAAGKLLALDSQLTNDPSEK 162
Db 249 NSADVVTREESDKFVRIDGLNATTEKLDTRLASAESKSIADHDTRLNGLDKTVSD----- 303
QY 163 SSYFQSQVDRIKEAYAG-----AAAGI-----VAGPFGLIISYSIAAGVIEGKLIPELN 212
Db 304 -----LRKETQGLAEQAALSGLPQPNVGGSGGVAADIGAGLADALTAP-LD 352
QY 213 DRLKAVQNFFTSLSVTYKQANKDIDAUKLATEIAA-----IGETKTETETTRF 262
Db 353 HKDKGLQ-----SLTLDQSVRKNE--KLKLAAGAEKTYGNGDSLTGKLNKD-KVSRP 403
QY 263 -----YVDYDDLMLSLKGAAKMINTCNEYQORHGKKTLLLEVPDI 303
Db 404 DFIQIEVDGQLITLESGEFQ-----VYKQSHSALTAFQTEQI 441

RESULT 40
US-10-220-481-155
; Sequence 155, Application US/10220481
; Publication No. US20040110670A1
; GENERAL INFORMATION:
; APPLICANT: ARICO, Maria B., et al.
; TITLE OF INVENTION: Heterologous Expression of Neisserial Proteins
; FILE REFERENCE: CHIR-15883/01US
; CURRENT APPLICATION NUMBER: US/10/220,481
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: PCT/IB01/00452
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 633
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 155
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 961c-741
; US-10-220-481-155

Query Match      6.8%; Score 102; DB 16; Length 590;
Best Local Similarity 21.4%; Pred. No. 5.7;
Matches 74; Conservative 62; Mismatches 129; Indels 80; Gaps 17;

QY 5 FAEQTV-----EVVKSIAETADGALDFYKYLDOVIPWKT-FDETIKELSRFKQEQSQ 56
Db 131 FAEETKTIVIKIDEKLEAVADTVDKHAEAFNDIADSLDETNTKAEAVKTANEAKQT-AE 189
QY 57 EASVLGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYI-----LLFDEYNEKKAAS 110
Db 190 ETKQNV-DAKVKAETAAGKAEAAAGTANTAADKAEAAKVTDIKADIATNKDNIKA 248
QY 111 AQKDIILIR-----ILDDGVNKLNEAOKSLGSSQ-SFNNAAGKLLALDSQLTNDPSEK 162
Db 249 NSADVVTREESDKFVRIDGLNATTEKLDTRLASAESKSIADHDTRLNGLDKTVSD----- 303
QY 163 SSYFQSQVDRIKEAYAG-----AAAGI-----VAGPFGLIISYSIAAGVIEGKLIPELN 212
```

```
; ORGANISM: Chlamydia trachomatis
; US-10-282-122A-55283

Query Match      6.8%; Score 102.5; DB 15; Length 472;
Best Local Similarity 23.1%; Pred. No. 3.8;
Matches 74; Conservative 42; Mismatches 105; Indels 99; Gaps 17;

QY 25 DFYKVLDOVIPWKTDEIK-----ELSRFKQEQSQASVLGDIKVLMLDS 72
Db 48 DFY--FLEHRIIFRVLDQAFKSDRPMDFLHTEELK-----RQQLNVIGGASYLI--- 96
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYIILFDEYNE-----KASAKDILILDL--- 121
Db 97 -----TLSEFAGT-----SAYI-----EEVADIIRSKSILRKMIOAAKDIEKKAABEP 139
QY 122 -DGVNKLNEAOKSLGSSQFNNA-----SGKLLALDSQLTNDP-----SEKSYFQSQVD 171
Db 140 RDVTTALDDAQNLLFRISQTTNFPVYVSDKLGKLSSTYKDKSFFLLALQERQEAQASH 199
QY 172 RIRKEAVAGAAGIV-----AGPFGLIISYSIAAGVIEGKLIPELNDRUKAVONF-- 221
Db 200 DARIPITLSGPTFLDLDRMLNGFSPSNLII-----LAARPAMGKTALALN-----IVENFCF 252
QY 222 -----FTLSVTYVKQ-----ANKDIDAUKLATEIA-----AIGEIKTETETTR 261
Db 253 ESRLPVGIFSLMTVDQLIHRITCSRSEVEAKKISVG-DISGRDFORVSVVREMEHTL 311
QY 262 FYVDYDDLMLSLKGAAKM 281
Db 312 LIDDYPLGLKITDLRARARM 331

RESULT 39
US-10-220-480-50
; Sequence 50, Application US/10220480
; Publication No. US20040092711A1
; GENERAL INFORMATION:
; APPLICANT: Chiron Spa
; TITLE OF INVENTION: Hybrid Expression of Neisserial Proteins
; FILE REFERENCE: P026783WO
; CURRENT APPLICATION NUMBER: US/10/220,480
; CURRENT FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: GB 0004695.3
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: GB 0027675.8
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 50
; LENGTH: 590
; TYPE: PRT
```

```
Db 304 -----LKEETQGLAEQAALSGLFQPVNCGSGGVAADIGAGLADALTAP-LD 352
Qy 213 DLKAVQNFFTSLSVTVKQAKDIDAALKLATEIAA-----ICEIKTETETTRF 262
Db 353 HKDKGLQ-----SLTLQSVRKNE--KLKLAQAQAEKTYGNGDSLNTGKLKND-KVSRF 403
Qy 263 ----VYDYLMLSLKGAQAKMINTCNEYQQRHGKKTLLLEVPDI 303
Db 404 DFIQIEVDGQLITLESGEFQ-----VYKQSHSALTAFQTEQI 441
```

RESULT 41

```
US-10-220-480-56
; Sequence 56, Application US/10220480
; Publication No. US20040092711A1
; GENERAL INFORMATION:
; APPLICANT: Chiron Spa
; TITLE OF INVENTION: Hybrid Expression of Neisserial Proteins
; FILE REFERENCE: P026783WO
; CURRENT APPLICATION NUMBER: US/10/220,480
; CURRENT FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: GB 0004695.3
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: GB 0027675.8
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 56
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 961cL-741
US-10-220-480-56
```

```
Query Match 6.8%; Score 102; DB 15; Length 612;
Best Local Similarity 21.4%; Pred. No. 6;
Matches 74; Conservative 62; Mismatches 129; Indels 80; Gaps 17;

Qy 5 FAEQTV-----EVVKSATETADGALDFYNYKLDQVIPWKT-FDETIKELSRPKQEQYSQ 56
Db 153 FAEETKTNIIVKIDKLEAVADVTDKHAFAFNDAIADSLDETNTKADAEVKTANEAKQT-AE 211
Qy 57 EASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYI-----LLFDEVNEKKAS 110
Db 212 ETQKQNV-DAKVKAAETAAGKAAEAGTANTAADKAAEVAARKVTDIKADIATNKONIAKKA 270
Qy 111 AQKDILIR-----ILDDGVNKLNEAOKSLIGSSQ-SFNNASGKLALDSQLTNDPSEK 162
Db 271 NSADVTTRESDSKFVRIDGLNATTEKLDTRLASAESAESIADHDTRLNGLDKTQVSD----- 325
Qy 163 SSYFQSDVDRIRKEAYAG-----AAAGI-----VAGPGLIISYSIAAGVIEGKLIPELN 212
Db 326 ----LRKETROGLAEQAALSGLFQPVNCGSGGVAADIGAGLADALTAP-LD 374
Qy 213 DLKAVQNFFTSLSVTVKQAKDIDAALKLATEIAA-----ICEIKTETETTRF 262
Db 326 ----LRKETROGLAEQAALSGLFQPVNCGSGGVAADIGAGLADALTAP-LD 374
Qy 163 SSYFQSDVDRIRKEAYAG-----AAAGI-----VAGPGLIISYSIAAGVIEGKLIPELN 212
Db 326 ----LRKETROGLAEQAALSGLFQPVNCGSGGVAADIGAGLADALTAP-LD 374
Qy 263 ----VYDYLMLSLKGAQAKMINTCNEYQQRHGKKTLLLEVPDI 303
Db 426 DFIQIEVDGQLITLESGEFQ-----VYKQSHSALTAFQTEQI 463
```

RESULT 42

```
US-10-220-481-161
; Sequence 161, Application US/10220481
; Publication No. US20040110670A1
; GENERAL INFORMATION:
; APPLICANT: ARICO, Maria B., et al.
; TITLE OF INVENTION: Heterologous Expression of Neisserial Proteins
; FILE REFERENCE: CHIR-15883/01US
```

```
; CURRENT APPLICATION NUMBER: US/10/220,481
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: PCT/IB01/00452
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 633
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 161
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 961cL-741
US-10-220-481-161
```

```
Query Match 6.8%; Score 102; DB 16; Length 612;
Best Local Similarity 21.4%; Pred. No. 6;
Matches 74; Conservative 62; Mismatches 129; Indels 80; Gaps 17;

Qy 5 FAEQTV-----EVVKSATETADGALDFYNYKLDQVIPWKT-FDETIKELSRPKQEQYSQ 56
Db 153 FAEETKTNIIVKIDKLEAVADVTDKHAFAFNDAIADSLDETNTKADAEVKTANEAKQT-AE 211
Qy 57 EASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYI-----LLFDEVNEKKAS 110
Db 212 ETQKQNV-DAKVKAAETAAGKAAEAGTANTAADKAAEVAARKVTDIKADIATNKONIAKKA 270
Qy 111 AQKDILIR-----ILDDGVNKLNEAOKSLIGSSQ-SFNNASGKLALDSQLTNDPSEK 162
Db 271 NSADVTTRESDSKFVRIDGLNATTEKLDTRLASAESAESIADHDTRLNGLDKTQVSD----- 325
Qy 163 SSYFQSDVDRIRKEAYAG-----AAAGI-----VAGPGLIISYSIAAGVIEGKLIPELN 212
Db 326 ----LRKETROGLAEQAALSGLFQPVNCGSGGVAADIGAGLADALTAP-LD 374
Qy 213 DLKAVQNFFTSLSVTVKQAKDIDAALKLATEIAA-----ICEIKTETETTRF 262
Db 375 HKDKGLQ-----SLTLQSVRKNE--KLKLAQAQAEKTYGNGDSLNTGKLKND-KVSRF 425
Qy 263 ----VYDYLMLSLKGAQAKMINTCNEYQQRHGKKTLLLEVPDI 303
Db 426 DFIQIEVDGQLITLESGEFQ-----VYKQSHSALTAFQTEQI 463
```

RESULT 43

```
US-10-766-993-3
; Sequence 3, Application US/10766993
; Publication No. US20050003510A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Chia-Hwa
; APPLICANT: Liu, Xiaowen
; APPLICANT: Lewicki, John A.
; APPLICANT: Xu, Qiang
; APPLICANT: Osel, Inc.
; TITLE OF INVENTION: Surface Expression of Biologically Active Proteins in
; FILE REFERENCE: Bacteria
; FILE REFERENCE: 016976-000810US
; CURRENT APPLICATION NUMBER: US/10/766,993
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US 60/443,619
; PRIOR FILING DATE: 2003-01-29
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1903
; TYPE: PRT
; ORGANISM: Lactobacillus jensenii 1153
; FEATURE:
; OTHER INFORMATION: genomic C370 cell wall anchor sequence
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1309)..(1903)
; OTHER INFORMATION: CWA200 cell wall associated region with anchor
; OTHER INFORMATION: motif
```

```
FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1868)..(1872)
; OTHER INFORMATION: anchor motif, sorting signal, cell wall targeting
; OTHER INFORMATION: region
US-10-766-993-3

Query Match      6.8%; Score 102; DB 17; Length 1903;
Best Local Similarity 19.1%; Pred. No. 28;
Matches 63; Conservative 42; Mismatches 109; Indels 116; Gaps 12;

QY 6 ABQTEVVVKSATETADGALD-----FYNKYLDQVVPWKT-----FDETTEL 47
DB 1225 SEEKQAFQKALDNKAKALDINSETEAEYKSADELQAKADLDGQTTDKSKLDDA 1284
QY 48 SRFK--QEYSQBASVLVGDIKVLLMDSQDKYFEATQTVYEWGVVTVQLLSAYILL 105
DB 1285 NNAKGTDKYKASD-----DTKSKFDALKA-----BEVK 1315
QY 106 EKASAKQDILIRILLDDGVNKLNEAKSILGSS-----QSFNNASGKLALD 155
DB 1316 NNSNATQKEV-----DDATNNLKQNNLNGQTTDKSKLDDAIDANNAGK- 1361
QY 156 TNDPSEKSYFQSVQDRIRKEAYAGAAAGIVAGPFLIISYSIAGVIEGKLIPEL 215
DB 1362 TDYKNASDDTKSKFDALKA-----BEVKNS 1390
QY 216 KAVQNFSTSLVTVVQAKNDIDAAKLKATETAAGIEIKTETETTRF----- 272
DB 1391 NATQKEVDATNNLKQAQNDLDGQ-----TTDKSKLDEAITDANNTKLTDK 1446
QY 273 LLKGAKKMINTCNEVQQRHGKTKLLEVPD 302
DB 1447 KFDEALKKAENKND-----SNATQKEVDD 1471

RESULT 44
US-10-032-585-7611
; Sequence 7611, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: BO, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Rusey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7611
; LENGTH: 1948
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7611

Query Match      6.8%; Score 102; DB 14; Length 1948;
Best Local Similarity 21.9%; Pred. No. 29;
Matches 60; Conservative 41; Mismatches 97; Indels 76; Gaps 11;

QY 25 DFYNKYLDQVVPWKTFTETIKELSRFKQEYSQBASVLVGDIKVLLMDSQDKYFE 81
DB 1434 DYIQLYQDIT---KTLKSTREELNGSKTEILR-----LKALLRESEDELYQ 1482
QY 82 --TVYEWGVVTVQL-----LSAYILLDFEYNEKKASAKOILIL--RIL 120
DB 1483 KTSVHDYEQDLAQLKVXKHTLLSRNKDINSELSIYKRSDEYTKLELAEISAISK 1542
QY 121 DDGVNKLNEAKSILGSSQSFNNASGKLALDQLNDPSEKSYFQSVQDRIRKEAY 180
DB 1543 EQATKEMKESRSQLLVRE-----LRTTQILIKDPRIKVENLEATIE----- 1586
```

```
QY 181 AAAGIVAGPFLIISYSIAGVIEGKLIPE--LNDRIKAVON--FFTSLSVTVQKANK 237
DB 1587 -----KNHOLDANKKEIKQIQDKLNYHLKNFENKELKEKKEIKNLRDLD 1633
QY 238 AAKLKLATEIAAIGEIKTETETTRFVVDYDDLML 271
DB 1634 -----FKTDI-----ETKLIKENKKQLQDYEDVILL 1658

RESULT 45
US-10-171-311-164
; Sequence 164, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatz, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-164

Query Match      6.7%; Score 101.5; DB 14; Length 1938;
Best Local Similarity 21.8%; Pred. No. 32;
Matches 69; Conservative 48; Mismatches 105; Indels 95; Gaps 14;

QY 7 EQTEVVVKSAT--ETAD-----GALDFYNKYLDQVVPWKTFTETIKELSRFK 51
DB 1173 EQEVTVLKALDEETRSHEAQVQEMRQKHAQAVEELTEQLEQ-----FKRAKANLDKNK 1226
QY 52 QEYSQBASVLVGDIKVLLMDSQDKYFEATQTVYEWGVVTVQLLSAYILLDFEYNEKKASA 111
DB 1227 QMLEKENADLAGELRVL-----GQAKQEV-----EHKKKLEA 1259
QY 112 QKDILIRILLDDG-----VNKLNEAKSILGSSQSFNNASGKLALDQLNDPSE 161
DB 1260 QVQELQSKCSQGERARAEKLVKLVKQNEVESVTG---MLNEAEGKAI-----KLAKDVAS 1312
QY 162 KSSYFQSQVDRIRKEAYAGAAAGIVAGPFLIISYSIAGVIE--GKLIPELNDRLKAV 218
DB 1313 LSSQLQDTQELLQEBETROK-----LNVSTKLQLEBEERNSLOQDLDEEMAK 1359
QY 219 QNF---FTSLSVTVQKANKDIDAAKLKATETETTRFVVDYDDLMLSLK 275
DB 1360 QNLEHSTLNIQSDSKKQLQ-----DPASTVEALEEGK-----KRFQKEIENLTQQVEE 1410
QY 276 GAA---KQMINTCNEYQQ 290
DB 1411 KAAAYDKLEKTKNRLQQ 1427

RESULT 46
US-09-927-597-2
```

; PRIOR APPLICATION NUMBER: US 60/335,936									
; PRIOR FILING DATE: 2001-11-14									
; NUMBER OF SEQ ID NOS: 238									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 162									
; LENGTH: 1972									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-171-311-162									
Query Match 6.7%; Score 101.5; DB 14; Length 1972;									
Best Local Similarity 21.8%; Pred. No. 33;									
Matches 69; Conservative 48; Mismatches 105; Indels 95; Gaps 14;									
QY	7	EQTVEVKS	AI--ETAD-----	GALDFY	NKYLDQ	VPKMTF	DETIKEL	SRFK	51
DB	1173	EQEVTVLK	ALDEETRS	HEAQVQ	EMRQKHAQ	AVEELTEQ	LEQ-----	FKR	AKANL
QY	52	QEYSQEA	SVLVDG	IKVLLM	DSQDKY	FEATQTV	YVCGVVT	QLLS	AYILLF
DB	1227	QTLEKENA	DLAGELRV	-----	GQAKQEV	-----	EHK	KKKLEA	1259
QY	112	QXDILIR	ILDDG-----	VNKLEA	QKSLG	SSQSFN	NASGKL	ALDLS	OLTND
DB	1260	QVQELSK	SCDGER	ARAE	LNLDKV	HKLQNE	VESVTG	---ML	NEAEGK
QY	162	KSSYFQ	SQVDRI	KAYAGA	AGIVAG	PFGLI	ISYSIA	AGVIE	---GKLI
DB	1313	LSSQLQ	TQELQ	QETTRQ	-----	LNVT	KLQLE	ERN	SLQDL
QY	219	QNF---	FTSL	SVTVK	QANKD	IDA	AKLATE	IAAIG	EITET
DB	1360	QNLERH	ISTLN	IQLS	DSKKLQ	----	DFAS	TV	EALEEG
QY	276	GAA--	KKMINT	CNEY	QO 290				
DB	1411	KAAAYD	KLEK	TNRL	QO 1427				
RESULT 48									
US-10-341-434-103									
; Sequence 103, Application US/10341434									
; Publication No. US20030215835A1									
; GENERAL INFORMATION:									
; APPLICANT: Origene Technologies									
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes									
; FILE REFERENCE: 9U 204 205 R1									
; CURRENT APPLICATION NUMBER: US/10/341,434									
; CURRENT FILING DATE: 2003-07-18									
; PRIOR APPLICATION NUMBER: US 60/348,164									
; PRIOR FILING DATE: 2002-01-15									
; PRIOR APPLICATION NUMBER: US 60/348,119									
; PRIOR FILING DATE: 2002-01-15									
; NUMBER OF SEQ ID NOS: 238									
; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO 103									
; LENGTH: 1972									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-341-434-103									
Query Match 6.7%; Score 101.5; DB 14; Length 1972;									
Best Local Similarity 21.8%; Pred. No. 33;									
Matches 69; Conservative 48; Mismatches 105; Indels 95; Gaps 14;									
QY	7	EQTVEVKS	AI--ETAD-----	GALDFY	NKYLDQ	VPKMTF	DETIKEL	SRFK	51
DB	1173	EQEVTVLK	ALDEETRS	HEAQVQ	EMRQKHAQ	AVEELTEQ	LEQ-----	FKR	AKANL
QY	52	QEYSQEA	SVLVDG	IKVLLM	DSQDKY	FEATQTV	YVCGVVT	QLLS	AYILLF
DB	1227	QTLEKENA	DLAGELRV	-----	GQAKQEV	-----	EHK	KKKLEA	1259

```
QY 112 QKDLIRILDG-----VNKLEAQKSLGSSQSFNNASGKLLALDSQLTNDPSE 161
Db 1260 QVQELQSKCSDGERARAEINLKVHKLQNEVESVTG---MLNEAEGKAI---KLAQDVAS 1312
QY 162 KSSYFQSDVRIRKEAYAGAAGIVAGPFGLLISYSIAAGVIE---GKLIPELNDRLKAV 218
Db 1313 LSSQLQDTQELLQEBTRQK-----LNVSTKLQLEERNSLQDLDEMEAK 1359
QY 219 QNF---FTSLSVTVKQANKDIDAAKLKLATEIAAIGETKTETTRFYVDYDDMLSLK 275
Db 1360 QNLERHISTLNIQLSDSKKLQ---DFASTVEALEEGK-----KRFQKEIENLTQOYEE 1410
QY 276 GAA---KMWINTCNEYQQ 290
Db 1411 KAAAYDKLEKTKNRLQQ 1427

RESULT 49
US-09-927-597-4
; Sequence 4, Application US/09927597
; Publication NO. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Pady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CYTOPO18
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-4

Query Match 6.7%; Score 101.5; DB 10; Length 1979;
Best Local Similarity 21.8%; Pred. No. 33;
Matches 69; Conservative 48; Mismatches 105; Indels 95; Gaps 14;

QY 7 EQTVVVVKSAI--ETAD-----GALDFYKNKYLDQVIPWKTFTDETIKELSRPK 51
Db 1180 EQBTVLTKALDETSHERAQVQEMRQKHAQVVEELTEQLQ-----FKRAKANLDRNK 1233
QY 52 QEYSQASVLVGDIKVLLMDSODKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASA 111
Db 1234 QTLKENADLAGELRVL-----GQAKQEV-----EHHKKKLEA 1266
QY 112 QKDLIRILDG-----VNKLEAQKSLGSSQSFNNASGKLLALDSQLTNDPSE 161
Db 1267 QVQELQSKCSDGERARAEINLKVHKLQNEVESVTG---MLNEAEGKAI---KLAQDVAS 1319
QY 162 KSSYFQSDVRIRKEAYAGAAGIVAGPFGLLISYSIAAGVIE---GKLIPELNDRLKAV 218
Db 1320 LSSQLQDTQELLQEBTRQK-----LNVSTKLQLEERNSLQDLDEMEAK 1366
QY 219 QNF---FTSLSVTVKQANKDIDAAKLKLATEIAAIGETKTETTRFYVDYDDMLSLK 275
Db 1367 QNLERHISTLNIQLSDSKKLQ---DFASTVEALEEGK-----KRFQKEIENLTQOYEE 1417
QY 276 GAA---KMWINTCNEYQQ 290
Db 1418 KAAAYDKLEKTKNRLQQ 1434
```

```
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6362
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6362

Query Match 6.7%; Score 101; DB 14; Length 1475;
Best Local Similarity 17.9%; Pred. No. 24;
Matches 53; Conservative 52; Mismatches 117; Indels 74; Gaps 8;

QY 7 EQTVVVVKSAI-----TADGALDFYKNKYLDQVIPWKTFTDETIKELSRPKQEYS 55
Db 231 EESIKLIQSQIELLRQSHSQSLSEIQEARIFEKMLTQQV-----DSAMKAKSDREAAK 285
QY 56 QEASVLVGDIKVLLMDSODKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKASAKQDI 115
Db 286 AREQVLEKQVQELRLKLEEDPEKNQLVHNLALNAQI-----EELTKAL 331
QY 116 LIRILDDGVNKLNEAQKSLGSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSDVRIRK 175
Db 332 KVDSMQOGATASEDRIRELIGGHQEAIKQLENTKOMNESLQDLVEKARFSEMERIRT 391
QY 176 EAYAGAAGIVAGPFGLLISY-----STAAGVIGKL-----IPELNDRL 215
Db 392 ESQTTSES-----LKYEHVLVKOMIEMDEKLEAEVLAKSQOANLEIQEPHDKI 441
QY 216 KAVQNFFTSLSVTVKQANKDIDAAKLKLATEIAA-----IGEIKTETTRTF 262
Db 442 KQLE-----LEVQLSSENKEKIQAEMLMVQEKASENIKNAEEKVNGLEAEVEKLRF 492

Search completed: January 28, 2005, 19:28:34
Job time : 148 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:58:34 ; Search time 62.7056 Seconds
(without alignments)
1749.716 Million cell updates/sec

Title: US-09-993-292B-2
Perfect score: 1515
Sequence: 1 MTSIFAQTVVVKSAIETA.....NEYQQRHGKTLFEPVDVAS 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/PCTUS_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1515	100.0	305	9	US-09-993-292A-2
2	130.5	8.6	889	16	US-10-437-963-197045
3	126	8.3	891	14	US-10-369-493-197045, A
4	116	7.7	1023	10	US-09-884-696-5
5	113.5	7.5	2823	14	US-10-369-493-5220
6	113.5	7.5	2823	14	US-10-369-493-5221
7	113	7.5	2805	14	US-10-369-493-1061
8	112.5	7.4	564	15	US-10-424-599-212070
9	112.5	7.4	1948	14	US-10-032-585-7611
10	112	7.4	1189	15	US-10-282-122A-70920
11	111.5	7.4	1158	15	US-10-282-122A-56861
12	111.5	7.4	1999	14	US-10-028-248A-107
13	111.5	7.4	1999	15	US-10-107-782-107

14	111	7.3	941	15	US-10-282-122A-46343	Sequence 46343, A
15	111	7.3	1295	14	US-10-369-493-6440	Sequence 6440, Ap
16	110	7.3	1938	14	US-10-369-493-5109	Sequence 5109, Ap
17	110	7.3	1938	14	US-10-369-493-5110	Sequence 5110, Ap
18	109.5	7.2	742	15	US-10-282-122A-53523	Sequence 53523, A
19	108.5	7.2	815	16	US-10-437-963-191043	Sequence 191043, A
20	108.5	7.2	3692	15	US-10-282-122A-71235	Sequence 71235, A
21	108	7.1	737	16	US-10-437-963-107270	Sequence 107270, A
22	107.5	7.1	1875	14	US-10-369-493-2285	Sequence 2285, A
23	106.5	7.0	344	14	US-10-369-493-497	Sequence 497, App
24	106.5	7.0	647	15	US-10-425-114-51665	Sequence 51665, A
25	106.5	7.0	857	14	US-10-369-493-9374	Sequence 9374, Ap
26	106.5	7.0	1312	14	US-10-393-602-148	Sequence 148, Appli
27	106	7.0	2835	9	US-09-885-535-4	Sequence 4, Appli
28	105.5	7.0	1318	16	US-10-408-765A-666	Sequence 666, App
29	105	6.9	990	15	US-10-108-260A-3896	Sequence 3896, Ap
30	105	6.9	1174	14	US-10-094-749-1911	Sequence 1911, Ap
31	104.5	6.9	584	17	US-10-425-115-226623	Sequence 226623, A
32	104.5	6.9	625	15	US-10-425-114-64494	Sequence 64494, A
33	104.5	6.9	2295	16	US-10-437-963-135452	Sequence 135452, A
34	104.5	6.9	3158	9	US-09-815-242-12611	Sequence 12611, A
35	104	6.9	529	15	US-10-282-122A-48338	Sequence 48338, A
36	104	6.9	809	15	US-10-282-122A-51808	Sequence 51808, A
37	104	6.9	1965	14	US-10-369-493-3279	Sequence 3279, Ap
38	103.5	6.8	568	17	US-10-425-115-267777	Sequence 267777, A
39	103.5	6.8	1583	16	US-10-408-765A-1635	Sequence 1635, Ap
40	103.5	6.8	1585	16	US-10-437-963-124349	Sequence 124349, A
41	103.5	6.8	1959	14	US-10-028-248A-36	Sequence 36, Appl
42	103.5	6.8	1959	15	US-10-107-782-36	Sequence 2687, Ap
43	103	6.8	5171	16	US-10-408-765A-2687	Sequence 2687, Ap
44	102.5	6.8	420	15	US-10-425-114-60225	Sequence 60225, A
45	102.5	6.8	957	16	US-10-437-963-176713	Sequence 176713, A
46	102.5	6.8	1961	14	US-10-028-248A-103	Sequence 103, App
47	102.5	6.8	1961	15	US-10-107-782-103	Sequence 103, App
48	102	6.7	451	14	US-10-369-493-2102	Sequence 2102, Ap
49	102	6.7	1015	14	US-10-369-493-3210	Sequence 3210, Ap
50	102	6.7	1038	14	US-10-032-585-7776	Sequence 7776, Ap

ALIGNMENTS

RESULT 1

US-09-993-292A-2
; Sequence 2, Application US/0993292A
; Patent No. US20020146430A1
; GENERAL INFORMATION:
; APPLICANT: James E. Galen
; APPLICANT: University of Maryland
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: UOPMD.007A
; CURRENT APPLICATION NUMBER: US/09/993,292A
; CURRENT FILING DATE: 2001-11-23
; PRIOR APPLICATION NUMBER: 60/252,516
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Salmonella Typhi
US-09-993-292A-2

Query Match 100.0%; Score 1515; DB 9; Length 305;
Best Local Similarity 100.0%; Pred. No. 9,9e-118;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLDQVTPKTFDETKLSRPFQKYSQASV 60

Db 1 MTSIFAQTVVVKSAIETADGALDLYNKYLDQVTPKTFDETKLSRPFQKYSQASV 60

QY 61 LVGDIKVLMDSQDKFEATQTVYGVGVVTVQLLSAYILLDFEYNEKKASQAKDILRIL 120

Wed Jan 5 14:16:58 2005

544 KSTKSEYVMDLEANYDITCLRKNVDKLEAVNKYRECESKET 587

Db 120
Qy 180
Db 180
Qy 240
Db 240
Qy 300
Db 300
Qy 301 PDVAS 305
Db 301 PDVAS 305

RESULT 2
US-10-437-963-197045
; Sequence 197045, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 197045
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_92840C.1.pap
US-10-437-963-197045

Query Match 8.6%; Score 130.5; DB 16; Length 889;
Best Local Similarity 22.1%; Pred. No. 0.068;
Matches 76; Conservative 66; Mismatches 123; Indels 79; Gaps 17;

Qy 54
Db 325
Qy 110
Db 384
Qy 166
Db 432
Qy 210
Db 486
Qy 257
Db 543
Qy 296

Db 120
Qy 180
Db 180
Qy 240
Db 240
Qy 300
Db 300
Qy 301 PDVAS 305
Db 301 PDVAS 305

RESULT 3
US-10-369-493-10918
; Sequence 10918, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10918
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Ferroplasma acidarmanus
US-10-369-493-10918

Query Match 8.3%; Score 126; DB 14; Length 891;
Best Local Similarity 22.8%; Pred. No. 0.16;
Matches 71; Conservative 52; Mismatches 119; Indels 70; Gaps 13;

Qy 561
Db 503
Qy 103
Db 614
Qy 163
Db 674
Qy 221
Db 706
Qy 278
Db 759

RESULT 4
US-09-884-696-5
; Sequence 5, Application US/09884696
; Publication No. US20030035809A1
; GENERAL INFORMATION:
; APPLICANT: GEORGE, LITTLE W
; APPLICANT: ANGELOS, JOHN A
; APPLICANT: HESS, JOHN F
; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES
; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA
; TITLE OF INVENTION: BOVIS INFECTIONS
; FILE REFERENCE: 481.06
; CURRENT APPLICATION NUMBER: US/09/884,696
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5


```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-07-27
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-03-35
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 70920
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70920

Query Match 7.4%; Score 112; DB 15; Length 1189;
Best Local Similarity 18.3%; Pred. No. 3.5;
Matches 61; Conservative 60; Mismatches 130; Indels 82; Gaps 9;

Qy 7 EQTEVVVK-----SATEADGALDLYNKYLDQVWPWKTDETIKELSRFKQYSQASV 60
Db 163 EESAGVLKYYKRAESQKLDHTEDNLNRVEDILYDLGRVPLKEEAIAKEYKLSKE 222
Qy 61 LVGDIKVLMDSDQKYFEATQTVYEWG-----GVVTTLSAYILLFDEYNE 106
Db 223 MEQSDVITVSDIDHYTDEQNRDLERLNLKSOAEKEGQQAQINQLQRY----- 273
Qy 107 KKAQAQKILIRIILDDGVKLNKAEQSLTSSOSFNASGKLLALDSQLTND-----FS 160
Db 274 -KGRQON-----DYDIEKLN---YELVKATENYEQLSGLNVLEERKKNQSETNARYE 323
Qy 161 EKSYFQSDVRIRKEAYAGAAAGVAGPGLIISYIAAGVIEGKLIPELNNRL-----K 216
Db 324 EELNLSQSDISIKNEKAQNE-----KLLADLKNKQKQLNK 359
Qy 217 TVQNFFTSLSATVQKANDDAKILATETAATGEIKTETTRTFYVDVDDLMLSLK 276
Db 360 EVOEESLLYVSDQHEKLEIKNSYITLMSQSDVNNDIRFLEHTINENAKSKRLDS 419
Qy 277 AAKMINTCNEYQ-----RHGKKTLEFV 300
Db 420 RLVEAFNLQKDIQNTQTNKAYQSSKSMQKV 452
```

```

RESULT 11
US-10-282-122A-56861
; Sequence 56861, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-07-27
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-03-35
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 56861
; LENGTH: 1158
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-10-282-122A-56861

Query Match 7.4%; Score 111.5; DB 15; Length 1158;
Best Local Similarity 21.0%; Pred. No. 3.7;
Matches 66; Conservative 58; Mismatches 121; Indels 69; Gaps 15;

Qy 4 IFAEQTEVVVKSAIETADGALDLYNKYLDQVWPWKTDETIKELSRFKQYSQASV 63
Db 183 VIAQONVN-----QITDTQL-LYQNQLEB-----TIQPPKPEIPMNYTTVLSTASILDE 230
Qy 64 DIKVLMDSDQKYFEATQTVYEWGVTQTLSSAVILLFDEYNKSKASQAQKILIRIILDDG 123
Db 231 DNK-LFTNEQEAHVKSQVT-----LLENSRSLESTSKTETDKSV-----DEY 274
Qy 124 VKKLINE-AQKSLTSSQSFNNSGKLLALDSQNTDNFSEKSSYFQSQVDRIRKEAYAGAA 182
Db 275 SKANEKIKKSIQAQNEQFERQK-----QELTNQWESDTTVYKQFQDLN----- 319
Qy 183 AGIVAGPFGLIISYS-----IAAGVIEGKLIPEL-NNRLKTVQNFFTSLSATVQKANK 234
Db 320 -GNVINQFSFYTPSQSGSIYADFLSESKLFQETQGNRIGELQKEIAELHMQVEQ--- 375
Qy 235 DIDAAKILKLA-----TEIAAIGEIKTETTRTFYVDVDDLMLSLKGAQKMWINTCNEYQ 290
```

Db 376 -LTALKQIAATYNDLEATPEIATDTQIKQ-----AIIQITNEKENIPNLDKNYQE 427

Qy 291 RHGKKTLEFEPDVA 304

Db 428 RL-EESLSEISFVS 440

RESULT 12

US-10-028-248A-107

; Sequence 107, Application US/10028248A

; Publication No. US20030235882A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard

; APPLICANT: Patturajan, Meera

; APPLICANT: Vernet, Corine

; APPLICANT: Casman, Stacie

; APPLICANT: Malyankar, Uriel

; APPLICANT: Shenoy, Suresh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gangolli, Esha

; APPLICANT: Miller, Charles

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Smithson, Glennda

; APPLICANT: Zethusen, Bryan

; APPLICANT: Liu, Xiaohong

; APPLICANT: Coleman, Steven

; APPLICANT: Tchernev, Velizar

; APPLICANT: Si, Jingsheng

; APPLICANT: Edinger, Shlomit

; APPLICANT: Stone, David

; APPLICANT: Sciore, Paul

; APPLICANT: Millet, Isabelle

; APPLICANT: Rothenberg, Mark

; TITLE OF INVENTION: NO. US20030235882A1el Nucleic Acids and Polypeptides and Methods

; TITLE OF INVENTION: Thereof

; FILE REFERENCE: 21402-222

; CURRENT APPLICATION NUMBER: US/10/028,248A

; CURRENT FILING DATE: 2001-12-19

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/256619

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/272408

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/285189

; PRIOR FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: 60/308039

; PRIOR FILING DATE: 2001-07-26

; PRIOR APPLICATION NUMBER: 60/311266

; PRIOR FILING DATE: 2001-08-09

; NUMBER OF SEQ ID NOS: 211

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 107

; LENGTH: 1999

; TYPE: PRT

; ORGANISM: Rattus norvegicus

US-10-028-248A-107

Query Match 7.4%; Score 111.5; DB 14; Length 1999;

Best Local Similarity 19.6%; Pred. No. 7.8;

Matches 60; Conservative 49; Mismatches 114; Indels 83; Gaps 11;

Qy 7 EGTVEVKSATP-----TADGALDLYNKYLDQVIP--WKTFTDET---IKELSRFKQEVQSQE 57

Db 1165 EQEVNLIKTBEEAKTHEAQIQEMRQKHSQAVEIAEQLEQTKRKVKANLEKAKOTLENE 1224

Qy 58 ASVLVDGIKVLMSDQKYFEATQTVYEWCGVVTQQLSAYILLPFDVYNKKASAKOILI 117

Db 1225 RGLANEVKKVLQGGED-----SEHKRKKVQAQLQELQ 1257

Qy 118 RILDDGVKLNKAQSKLLTSSQSFNNASCKLLALD---SOLTNDSEKSSYFQSQVDRI 174

Db 1258 VKFNEGERVTLADKVKTLQVELDNVTGLLSQSDSKSKLTKDFSALQSQDQTQELLQ 1317

Qy 175 KEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIIPELNNRLKTQV---NFTTSLSATVK 230

Db 1318 EEN-----RQKL--SLSTKLAQVEDEKNSFREQLSEEE 1349

Qy 231 QANKDIDAARKLATEIAAIGIKTETTRFYVDYDDL-MLSLKGAAKMINTCNEYQ 289

Db 1350 EAKHNLEK---QIATLHAQVADMKKME-----DSVGCLETAEEVKRKLQKQDLEGLS 1398

Qy 290 QRHGKK 295

Db 1399 QRHEEK 1404

RESULT 13

US-10-107-782-107

; Sequence 107, Application US/10107782

; Publication No. US20040018970A1

; GENERAL INFORMATION:

; APPLICANT: Boldog, Ferenc

; APPLICANT: Casman, Stacie

; APPLICANT: Coleman, Steve

; APPLICANT: Edinger, Shlomit

; APPLICANT: Gangolli, Esha

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Li, Li

; APPLICANT: Liu, Xiaohong

; APPLICANT: Malyankar, Uriel

; APPLICANT: Miller, Charles

; APPLICANT: Millet, Isabelle

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Sciore, Paul

; APPLICANT: Shenoy, Suresh

; APPLICANT: Shimkets, Richard

; APPLICANT: Si, Jingsheng

; APPLICANT: Smithson, Glennda

; APPLICANT: Spytek, Kimberly

; APPLICANT: Stone, David

; APPLICANT: Taupier, Raymond, jr.,

; APPLICANT: Tchernev, Velizar

; APPLICANT: Zethusen, Brian

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF

; FILE REFERENCE: 21402-222CIP

; CURRENT APPLICATION NUMBER: US/10/107,782

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: 10/028,248

; PRIOR FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/256,619

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/262,959

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/272,408

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/285,189

; PRIOR FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: 60/308,039

; PRIOR FILING DATE: 2001-07-26

; PRIOR APPLICATION NUMBER: 60/311,266

; PRIOR FILING DATE: 2001-08-09

; PRIOR APPLICATION NUMBER: 60/279,344

; PRIOR FILING DATE: 2001-03-28

; NUMBER OF SEQ ID NOS: 215

; SOFTWARE: Curaseqlist version 0.1

; SEQ ID NO 107

; LENGTH: 1999

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-107-782-107

Db 875 ALASQL-EEI-QHKLUVGESQVENKEEL-----ICAKIMNKEMVD 913
Qy 210 ELNNRLKTVQNFFTSLSATVKQAKOIDAAKKLATEIA-----AIGEIKT--- 255
Db 914 ELNAKLGDALEGMELKKSLEVSSEAKVORREBELIAQVSKHRDQEQEQLQTLDELKSAQH 973
Qy 256 ETETTR 261
Db 974 STETSR 979

RESULT 16
US-10-369-493-5109
; Sequence 5109, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5109
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5109

Query Match 7.3%; Score 110; DB 14; Length 1938;
Best Local Similarity 18.9%; Pred. No. 9.9;
Matches 55; Conservative 60; Mismatches 114; Indels 62; Gaps 9;
Qy 5 FAEQTVVVK-----SAIETADGALDLYNKYLQVDPKTFDETIKELSRFK 51
Db 1315 FSSQLVEAKAAEDLHERQEFHACKNLEHLDQCHELLEEQINGK--DDIQRLSRIN 1372
Qy 52 QEYSQ-----EASVLVG--DIKVLMDSQKYFEATQTVYEWCGVVTTOLLSAYILLFDEY 104
Db 1373 SEISQWKARYEGEGLVGSLEELKQKQNRVMDLQEAALSAQNKVILEKAKGKLLAET 1432
Qy 105 NEKASAKDILI-----RIIDDGVKKLNEAKSLLTSSQSFNNAAGKLLALD 152
Db 1433 EDARSVDRLHTVIAASLEKKQKQAFDKIVDDWKRVDDIQKEIDATTRDSRNTSTEVFKLR 1492
Qy 153 SOLTNDPSEKSYFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYSIAAGVIE-----GKL 207
Db 1493 SSMDN-----LSQIETLARE-----NKIFSQEI-RDINEQITQGGRT 1529
Qy 208 IPENLNRLKTVQNFFTSLSATVKQAKOIDAAKKLATEIAAIGIKTETE 258
Db 1530 YQEVHKSVRRLQEQKDELQHALDEAFALEABESKVLRLQIEVQVQIRSEIE 1580

RESULT 17
US-10-369-493-5110
; Sequence 5110, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5109
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5109

FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 5110
LENGTH: 1938
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-369-493-5110

Query Match 7.3%; Score 110; DB 14; Length 1938;
Best Local Similarity 18.9%; Pred. No. 9.9;
Matches 55; Conservative 60; Mismatches 114; Indels 62; Gaps 9;
Qy 5 FAEQTVVVK-----SAIETADGALDLYNKYLQVDPKTFDETIKELSRFK 51
Db 1315 FSSQLVEAKAAEDLHERQEFHACKNLEHLDQCHELLEEQINGK--DDIQRLSRIN 1372
Qy 52 QEYSQ-----EASVLVG--DIKVLMDSQKYFEATQTVYEWCGVVTTOLLSAYILLFDEY 104
Db 1373 SEISQWKARYEGEGLVGSLEELKQKQNRVMDLQEAALSAQNKVILEKAKGKLLAET 1432
Qy 105 NEKASAKDILI-----RIIDDGVKKLNEAKSLLTSSQSFNNAAGKLLALD 152
Db 1433 EDARSVDRLHTVIAASLEKKQKQAFDKIVDDWKRVDDIQKEIDATTRDSRNTSTEVFKLR 1492
Qy 153 SOLTNDPSEKSYFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYSIAAGVIE-----GKL 207
Db 1493 SSMDN-----LSQIETLARE-----NKIFSQEI-RDINEQITQGGRT 1529
Qy 208 IPENLNRLKTVQNFFTSLSATVKQAKOIDAAKKLATEIAAIGIKTETE 258
Db 1530 YQEVHKSVRRLQEQKDELQHALDEAFALEABESKVLRLQIEVQVQIRSEIE 1580

RESULT 18
US-10-282-122A-53523
; Sequence 53523, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

Query Match 7.2%; Score 109.5; DB 15; Length 742;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 66; Conservative 49; Mismatches 131; Indels 63; Gaps 10;

QY 18 ETADGALDLYNKVDQVWPKTFDETIKELSPKQYQEAQVGVGDIKVLMSQDKYF 77
DB 236 KTVDLASDATSKLSDTVQDKISDLPTIKT-----LNDTKLSSDLKFLFEDTNDLND 288
QY 78 EATQTVVWGVVTVQLLSAVILLFDBYNEKASAKDILIRILD-----DGVKKLNEAQ 131
DB 289 ELSPLIKSILNLMVDL-----SSASSLTNLIDAVNGSGEDVFKLIDNL 333
QY 132 KSLTSSQSFNNASGKLLALDSQLTN-----DFSEKSSYFQSQVDRIRKEAYA 179
DB 334 SEKLSNLQSLNDLVDFLTKLQNTSNRLDDVIDNLEDSSNKIDSSISTLNDIKNKVIS 393
QY 180 GAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRL-KTVQNFFTLSATVKQANKOI-- 236
DB 394 GQQPSISA--LNNVLSLNGICRINLNLNPNFDSKISKPINNIFAN---SIKVANDIITV 448
QY 237 -DRAKLG-----LATEATAEIKETETTFRFYVDYDMLSLKGAAKMNTCNEY 288
DB 449 LDKAEAKLPKEVIELTTSKLSGNAQESISLIR-----ERLPLAKGMLDDLIDTUSKI 501
QY 289 QQRHGKKTLL 297
DB 502 SNGEDMKKL 510

Query Match 7.2%; Score 108.5; DB 16; Length 815;
Best Local Similarity 20.5%; Pred. No. 4.1;
Matches 71; Conservative 69; Mismatches 121; Indels 85; Gaps 16;

QY 5 FAEQTVVVKSAIETADGALDLYNKVDQVWPKTFDETIKELSRPKQYQEAQVGVGDIKVLMSQDKYF 316
DB 261 FQEQEASLLDQLSSTKTKVTLSSEFRRE---KTAEELREIRRLRLESSLAQAGDDKDV 316
QY 61 L-----VGDIKVLMSQDKYFEATQTVVWGVV-----VTQLLSA-----YILLFDEYN 105
DB 317 LEAKLKEKGLDGNIL-----QEKVSLLSQEI-DNKGIRIRELSSLLSSKEADYRNLCSFSD 371
QY 106 EKKASQAQ-KDILIRILDGKVLKNEBAQSKLLTS-----SQSPFNASGKLLALDSQLTN 157
DB 372 QTKESLELAQAKIQLEBEVHRTRNDLSSKISSIDLLNEELQALNSAKNEAEKLSLTK 431
QY 158 DFSEKSSYFQSQVDRIRKEAYAGAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRLKT 217
DB 432 DYTDLKASSEARENSRSE-----LLLEKDNMKIQLDGLKSLDALSDDSSKD 475
QY 218 VQNFFTLSATVKQANKDIDDAKLAETIAAIGIK-----TETETTRFVVD 265
DB 476 REN-----TAAINKELDATKAMLENAVAVKSLRESLOSTBEALTDSRSEVSKLSVE 527
QY 266 YD-----DLMLSLKGAAKMNTCNEYQQRHGKKTLLFEVPDVA 304
DB 528 LDEANRMNQDLVLIQI-----SKLQDEFNEMQEGLTNK-LGEVESVS 567

Query Match 7.2%; Score 109.5; DB 15; Length 742;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 66; Conservative 49; Mismatches 131; Indels 63; Gaps 10;

QY 18 ETADGALDLYNKVDQVWPKTFDETIKELSPKQYQEAQVGVGDIKVLMSQDKYF 77
DB 236 KTVDLASDATSKLSDTVQDKISDLPTIKT-----LNDTKLSSDLKFLFEDTNDLND 288
QY 78 EATQTVVWGVVTVQLLSAVILLFDBYNEKASAKDILIRILD-----DGVKKLNEAQ 131
DB 289 ELSPLIKSILNLMVDL-----SSASSLTNLIDAVNGSGEDVFKLIDNL 333
QY 132 KSLTSSQSFNNASGKLLALDSQLTN-----DFSEKSSYFQSQVDRIRKEAYA 179
DB 334 SEKLSNLQSLNDLVDFLTKLQNTSNRLDDVIDNLEDSSNKIDSSISTLNDIKNKVIS 393
QY 180 GAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRL-KTVQNFFTLSATVKQANKOI-- 236
DB 394 GQQPSISA--LNNVLSLNGICRINLNLNPNFDSKISKPINNIFAN---SIKVANDIITV 448
QY 237 -DRAKLG-----LATEATAEIKETETTFRFYVDYDMLSLKGAAKMNTCNEY 288
DB 449 LDKAEAKLPKEVIELTTSKLSGNAQESISLIR-----ERLPLAKGMLDDLIDTUSKI 501
QY 289 QQRHGKKTLL 297
DB 502 SNGEDMKKL 510

Query Match 7.2%; Score 108.5; DB 16; Length 815;
Best Local Similarity 20.5%; Pred. No. 4.1;
Matches 71; Conservative 69; Mismatches 121; Indels 85; Gaps 16;

QY 5 FAEQTVVVKSAIETADGALDLYNKVDQVWPKTFDETIKELSRPKQYQEAQVGVGDIKVLMSQDKYF 316
DB 261 FQEQEASLLDQLSSTKTKVTLSSEFRRE---KTAEELREIRRLRLESSLAQAGDDKDV 316
QY 61 L-----VGDIKVLMSQDKYFEATQTVVWGVV-----VTQLLSA-----YILLFDEYN 105
DB 317 LEAKLKEKGLDGNIL-----QEKVSLLSQEI-DNKGIRIRELSSLLSSKEADYRNLCSFSD 371
QY 106 EKKASQAQ-KDILIRILDGKVLKNEBAQSKLLTS-----SQSPFNASGKLLALDSQLTN 157
DB 372 QTKESLELAQAKIQLEBEVHRTRNDLSSKISSIDLLNEELQALNSAKNEAEKLSLTK 431
QY 158 DFSEKSSYFQSQVDRIRKEAYAGAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRLKT 217
DB 432 DYTDLKASSEARENSRSE-----LLLEKDNMKIQLDGLKSLDALSDDSSKD 475
QY 218 VQNFFTLSATVKQANKDIDDAKLAETIAAIGIK-----TETETTRFVVD 265
DB 476 REN-----TAAINKELDATKAMLENAVAVKSLRESLOSTBEALTDSRSEVSKLSVE 527
QY 266 YD-----DLMLSLKGAAKMNTCNEYQQRHGKKTLLFEVPDVA 304
DB 528 LDEANRMNQDLVLIQI-----SKLQDEFNEMQEGLTNK-LGEVESVS 567

Query Match 7.2%; Score 109.5; DB 15; Length 742;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 66; Conservative 49; Mismatches 131; Indels 63; Gaps 10;

QY 18 ETADGALDLYNKVDQVWPKTFDETIKELSPKQYQEAQVGVGDIKVLMSQDKYF 77
DB 236 KTVDLASDATSKLSDTVQDKISDLPTIKT-----LNDTKLSSDLKFLFEDTNDLND 288
QY 78 EATQTVVWGVVTVQLLSAVILLFDBYNEKASAKDILIRILD-----DGVKKLNEAQ 131
DB 289 ELSPLIKSILNLMVDL-----SSASSLTNLIDAVNGSGEDVFKLIDNL 333
QY 132 KSLTSSQSFNNASGKLLALDSQLTN-----DFSEKSSYFQSQVDRIRKEAYA 179
DB 334 SEKLSNLQSLNDLVDFLTKLQNTSNRLDDVIDNLEDSSNKIDSSISTLNDIKNKVIS 393
QY 180 GAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRL-KTVQNFFTLSATVKQANKOI-- 236
DB 394 GQQPSISA--LNNVLSLNGICRINLNLNPNFDSKISKPINNIFAN---SIKVANDIITV 448
QY 237 -DRAKLG-----LATEATAEIKETETTFRFYVDYDMLSLKGAAKMNTCNEY 288
DB 449 LDKAEAKLPKEVIELTTSKLSGNAQESISLIR-----ERLPLAKGMLDDLIDTUSKI 501
QY 289 QQRHGKKTLL 297
DB 502 SNGEDMKKL 510

Query Match 7.2%; Score 108.5; DB 16; Length 815;
Best Local Similarity 20.5%; Pred. No. 4.1;
Matches 71; Conservative 69; Mismatches 121; Indels 85; Gaps 16;

QY 5 FAEQTVVVKSAIETADGALDLYNKVDQVWPKTFDETIKELSRPKQYQEAQVGVGDIKVLMSQDKYF 316
DB 261 FQEQEASLLDQLSSTKTKVTLSSEFRRE---KTAEELREIRRLRLESSLAQAGDDKDV 316
QY 61 L-----VGDIKVLMSQDKYFEATQTVVWGVV-----VTQLLSA-----YILLFDEYN 105
DB 317 LEAKLKEKGLDGNIL-----QEKVSLLSQEI-DNKGIRIRELSSLLSSKEADYRNLCSFSD 371
QY 106 EKKASQAQ-KDILIRILDGKVLKNEBAQSKLLTS-----SQSPFNASGKLLALDSQLTN 157
DB 372 QTKESLELAQAKIQLEBEVHRTRNDLSSKISSIDLLNEELQALNSAKNEAEKLSLTK 431
QY 158 DFSEKSSYFQSQVDRIRKEAYAGAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRLKT 217
DB 432 DYTDLKASSEARENSRSE-----LLLEKDNMKIQLDGLKSLDALSDDSSKD 475
QY 218 VQNFFTLSATVKQANKDIDDAKLAETIAAIGIK-----TETETTRFVVD 265
DB 476 REN-----TAAINKELDATKAMLENAVAVKSLRESLOSTBEALTDSRSEVSKLSVE 527
QY 266 YD-----DLMLSLKGAAKMNTCNEYQQRHGKKTLLFEVPDVA 304
DB 528 LDEANRMNQDLVLIQI-----SKLQDEFNEMQEGLTNK-LGEVESVS 567

Query Match 7.2%; Score 109.5; DB 15; Length 742;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 66; Conservative 49; Mismatches 131; Indels 63; Gaps 10;

QY 18 ETADGALDLYNKVDQVWPKTFDETIKELSPKQYQEAQVGVGDIKVLMSQDKYF 77
DB 236 KTVDLASDATSKLSDTVQDKISDLPTIKT-----LNDTKLSSDLKFLFEDTNDLND 288
QY 78 EATQTVVWGVVTVQLLSAVILLFDBYNEKASAKDILIRILD-----DGVKKLNEAQ 131
DB 289 ELSPLIKSILNLMVDL-----SSASSLTNLIDAVNGSGEDVFKLIDNL 333
QY 132 KSLTSSQSFNNASGKLLALDSQLTN-----DFSEKSSYFQSQVDRIRKEAYA 179
DB 334 SEKLSNLQSLNDLVDFLTKLQNTSNRLDDVIDNLEDSSNKIDSSISTLNDIKNKVIS 393
QY 180 GAAAGIVAGPGLIISYSTAAGVIEGKLIPELNNRL-KTVQNFFTLSATVKQANKOI-- 236
DB 394 GQQPSISA--LNNVLSLNGICRINLNLNPNFDSKISKPINNIFAN---SIKVANDIITV 448
QY 237 -DRAKLG-----LATEATAEIKETETTFRFYVDYDMLSLKGAAKMNTCNEY 288
DB 449 LDKAEAKLPKEVIELTTSKLSGNAQESISLIR-----ERLPLAKGMLDDLIDTUSKI 501
QY 289 QQRHGKKTLL 297
DB 502 SNGEDMKKL 510

Query Match 7.2%; Score 108.5; DB 16; Length 815;
Best Local Similarity 20.5%; Pred. No. 4.1;
Matches 71; Conservative 69; Mismatches 121; Indels 85; Gaps 16;

QY 5 FAEQTVVVKSAIETADGALDLYNKVDQVWPKTFDETIKELSRPKQYQEAQVGVGDIKVLMSQDKYF 316
DB 261 FQEQEASLLDQLSSTKTKVTLSSEFRRE---KTAEELREIRRLRLESSLAQAGDDKDV 316

FEATURE: Clone ID: PAT MET4530 8739C.1 den
OTHER INFORMATION:
PRIORITY: 1
REMARKS: Remaining Prior Application data removed - See File Wrapper or PALM.
PRIOR FILING DATE: 2001-02-16

RESULT 23
US-10-369-493-497 ; Sequence 497, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 497
LENGTH: 344
TYPE: PRT
ORGANISM: Xenorhabdus nematophilus
US-10-369-493-497

Query Match 7.0%; Score 106.5; DB 14; Length 344;
Best Local Similarity 21.2%; Pred. No. 1.9;
Matches 56; Conservative 52; Mismatches 103; Indels 53; Gaps 13;
QY 15 SAETADGALDLYNKYLQVDPKWT--FDETIKELSRFKQEQSQAQKDI--LIRILDGKVLMD 72
DB 71 TVLHVGDSDIRLPHQDISLPPKELGIDIVLDTGKYGERIDGEGAHLAGAKKVLFAHP 130
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQKDI--LIRILDGKVLMD 129
DB 131 GGNLDLAT-VY--GVNHLLSA-----EDCIVSNASCTNCIPIIKVLDD----- 174
QY 130 AQKSLTSSQFNAGVIEGLIPELNNRLKTQVNFSTLSATVKQANKOIDAAKCLKLATEIAA 249
DB 175 -----AFNIESGTVTTHASM-NDQPVIDAY-HSDLRTR-----AASQSIIP-- 215
QY 190 FGLIISYIAAGVIEGLIPELNNRLKTQVNFSTLSATVKQANKOIDAAKCLKLATEIAA 249
DB 216 -----VDTKLAAGIT--RIPPKFNRPFAISVRVPTINVTADILSVTSA-----AVNVSE 264
QY 250 IGBI--KTETETTFRFYVDYDML 271
DB 265 VNQLQKSAVSPRGIVDFDLPL 288

RESULT 24
US-10-425-114-51665
Sequence 51665, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 51665
LENGTH: 647
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700221631_FLI.pep
US-10-425-114-51665

Query Match 7.0%; Score 106.5; DB 15; Length 647;
Best Local Similarity 22.8%; Pred. No. 4.4;
Matches 79; Conservative 52; Mismatches 121; Indels 95; Gaps 15;
QY 9 TVEVKSATADGALDLYNKYLQVDPKWTFTETIKELSRFKQEQSQAQKDI--LIRILDGKVL 68
DB 79 TVQLEKARYELANADAKNEALSQV-----DDAV-----RASEAKTQVEHLIAVTHL 128

QY 69 --LWDS-----QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQKDI--LIRIL 120
DB 129 KGLVDSKVDGNRKKTATRIQNLEE-----ENFALKLELOKKAABEKAIRLERMI 178
QY 121 DD-----GVK-----KLENAQKSLTSSQFNAGVIEGLIPELNNRLKTQVNFSTLSATVKQANKOIDAAKCLKLATEIAA 152
DB 179 DELQSDVDDARNFGSKQLADEWQKKAQLLEVRLEADQSNILKGSLSNASEELDSTS 238
QY 153 SOLTNDPSE-----KSSYFQSQVDRIKRAY-----AGAAAGIVAGPFGLLIISYSTAAG 201
DB 239 SLLRDRESEVAALRDVRFLEDELTKLNDIVVSGORADAAEKAADLWTEVEGLRLKLH 298
QY 202 VIEGKLIPELNNRLKTQVNFSTLSATVKQANKOIDAAKCLKLATEIAA 254
DB 299 TVSEKIEALNSDKSTLSLETETNEOKNQNLADKLEANKD-ELEKTKAMEGLASALQEMS 357
QY 255 TET-----ETTRFYVDYDMLSLKGLAAGKAMINTCNEYQ 289
DB 358 AESREAREKYLKQDIERAQAQVEELSLSLK-----NTKENYE 396

RESULT 25
US-10-369-493-9374
Sequence 9374, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 9374
LENGTH: 857
TYPE: PRT
ORGANISM: Xylella fastidiosa
US-10-369-493-9374

Query Match 7.0%; Score 106.5; DB 14; Length 857;
Best Local Similarity 24.3%; Pred. No. 6.4;
Matches 69; Conservative 45; Mismatches 103; Indels 67; Gaps 14;
QY 7 EQTVVVKSAIETADGALDLYNK-----YLDQVIFPKTFDETIKELSRFKQEQSQAQKDI--LIRIL 60
DB 574 EEAIVKVSADARRSRTGLSDPNRPSGFLPGTGVGKT--ELCKALAEFLD-SODAMV 630
QY 61 LVGDIKVLMDSDQKYFEAT-----QTVYEWCGVVTQLLSA--VILLFDEYNEKKA 110
DB 631 -----RIDMSEFWEKHSVARLIGAPGVYEGGVTLELVRPRLSLILLDEV-EKAHS 684
QY 111 AQXDILIRILDGKVLKLENAQKSLTSSQFNAGVIEGLIPELNNRLKTQVNFSTLSATVK 170
DB 685 DVFNILLQVLDDG--RLTDQGRV---DPRNT---VIVMTSLGSHQIQLSGDDSD-- 733
QY 171 DRIKRAYAGAAAGIVAGPFGLLIISYIAAGVIEGLIPELNNRLKTQVNFSTLSATVK 230
DB 734 ---PEVYTQMKAAVM-----GVYQAHFRPEFIRLDDIVVFHPLDKAQIK 775
QY 231 QANKOIDAAKCLKLATEIAAIGETIKTETETTRFYVDYDMLSLK 274
DB 776 Q-----IARQLRGLERKLAESEUK-----LDLDDRALELL 806

RESULT 26
US-10-393-602-148
Sequence 148, Application US/10393602

Publication No. US20030170714A1
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/393,602
FILING DATE: 19-Mar-2003
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-393-602-148
Query Match 7.0%; Score 106.5; DB 14; Length 1312;
Best Local Similarity 18.9%; Pred. No. 11;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
QY 10 VEVKSAIETADGALDLYNKYLDQVVPKTFDETIK-----ELSRFKQRY 54
DB 502 MEVISIQNEKAD--LDRTLRKLDQMEQLNHHHTTTRTQMEMLTKDKADKDEQIRKIKSRH 559
QY 55 SEASVILVGDIVKLLMDSQDKYFEATQTVYVCGVVTQLLSAYILLFDEYNEKKAQAQD 114
DB 560 SDELTSLG-----YFPNKKQLEDWLHKSQ-----EINQTR 591
QY 115 ILIRILDGDKVKKLNEAKQSILTSQSQFNNAAGLSDQLTNDPSEK-----SSYFQS 168
DB 592 -----DLAKLN-----KELASSEQNKNHNNELKREBQLSS-YEDKLFVCGSQDFES 640
QY 169 QVDRIRKE-----AVAGAAAGIVAGPF----- 190
DB 641 DLDRLEEIEKSKORAMLAGATA--VYSQFTQLTDENOSCPVCQVQRFQTEAEQLQEV 698
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217
DB 699 SDIQSKLRAPDKLSTESLEKKEKRRDEMLGLVPMRQSIID--LKEKEIPELNNKLQN 756
QY 218 VQNFFTSLSATVQAANKNDAAKLKLAETAAIGIKTTETTT-----RPYVDY 266
DB 757 V-----NRDIQRLKNDIBEQETLLGTIMPEESAKVCLDVTIMERFQML 802
QY 267 DDMLSLKGAAK-----KMINTCNEYQQRHGK 295
Db 803 KDVERKTAQQAQAKLQGGIDLDRTVQOVNQEKQKH 838
RESULT 27
US-09-885-535-4
; Sequence 4, Application US/09885535
; Patent No. US20020104105A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-266-II
; CURRENT APPLICATION NUMBER: US/09/885,535
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,245
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 2835
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-885-535-4
Query Match 7.0%; Score 106; DB 9; Length 2835;
Best Local Similarity 19.4%; Pred. No. 36;
Matches 74; Conservative 76; Mismatches 122; Indels 110; Gaps 17;
QY 7 EQTVVVK-----SAIETAD-----GALDLYNKYLDQV-IPWKT 39
DB 807 KSTVEIKREGEKIAATPAADKVKILKQLSLDSRWEALLNKATRNQLEGISVVAQQ 866
QY 40 FDETIKELSRFK-----QYSQSEASVL---VGDIKVLLMDSQDKYFEATQTVYEW 86
DB 867 FHETLEPLNEWLTTEKRLVNCPEPIGTQASKLEEQIAQHKVQLQEDILLRKQNVQDALLNG 926
QY 87 CGVVTQLLSAYILLFDEYNEKKAQAQD-----LIRILDGV---KKLNEAKQSILTS 137
DB 927 LELLKQTTGDEVLIIOQKLEAIKARYKDITKLTSTDAKTLEQALQALRLHSTHELCT- 985
QY 138 SQSFNNAAGLJALDSQLTNDPSEKSSYFQSQVDRIKREYAGAAAGIVAGPFGLIISYS 197
DB 986 --WLDKVEVELLSYETQVLK--GEASQAQMRPKELKKEAKNKA-----LIDSLN 1032
QY 198 -IAAGVIE-----GKLIPELNNRLKTQNFFTSLSATVQAANKDIDA----- 238
DB 1033 EVSSALLELVPWRAREGLEKMAEDNERYLV-----SDTITQKVEIDAILRSQQF 1085
QY 239 -----AKKLATE-----IAAGIKTETETTFYVD-----YDDLMLSLKGA 278
DB 1086 DQADAELSMTTEKIMSLGDIRLEQDQTSALQVQKTFMTEILRHKDIIDDLVK-SG 1144
QY 279 KMINTCNEYQQRHGKTLFEV 300
DB 1145 HKINTACSEEEKQSMKKLDKV 1166
RESULT 28
US-10-408-765A-666
; Sequence 666, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Watnook, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

```
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 666
; LENGTH: 1318
; TYPE: PRP
; ORGANISM: Homo sapiens
; US-10-408-765A-666

Query Match      7.0%; Score 105.5; DB 16; Length 1318;
Best Local Similarity 18.4%; Pred. No. 14;
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;

QY 10 VEVVKSATETADGALDLYNKYLDQVTPWKTETETIK-----ELSRFQKEY 54
DB 508 MEVISLQNEKAD--LDRTLRKLDQEMEQNLNHTTTRTQMEMLTKDKADKDEQIRKIKSRH 565
QY 55 SQEASVLVGDIVKLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKQD 114
DB 566 SDELTSLG-----YFNKKQLEDWLHLSK-----EINQTR----- 597
QY 115 ILRILDDGVKLLNEAOKSLTSSQFNNSAGKLLALDLSQTLNDFSEKSSYFQSQVDR 172
DB 598 -----DRLAKLN---KELASQNKHNINNELERKEEQUSS-YEDKLFVCGSQDFES 646
QY 169 QVDRIRKE-----AYAGAAA-----GIVAGPFGLIISYSTAAGV 202
DB 647 DLDRLEKEIEKSKORAWLAGATVYSQFITQLTDENQSCPCVQCFQTEAEIQAISD 706
QY 203 IEGKL-----IPELNNRLKTVQNF 221
DB 707 LQSKLRAPDKLJSTSELKKKKRDEMLGLAPMQSIIDLKEKEIPELRNKLQNV--- 763
QY 222 FTSLSATVQKQNDIDAALKLATEIAAIGETETETTT-----RFVVDYDLM 270
DB 764 -----NRDIOBLKNDIEQETLLGTIMPEESAKVCLTDTVIMERFQMEKLDVE 812
QY 271 LSLKGAAG-----KWMINTCNEYQORHGKK 295
DB 813 RKIAQQAQKLOGIDLDRTVQVQVQKQKQKH 844

RESULT 29
US-10-108-260A-3896
; Sequence 3896, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3896
; LENGTH: 990
; TYPE: PRP
; ORGANISM: Homo sapiens
; US-10-108-260A-3896

Query Match      6.9%; Score 105; DB 15; Length 990;
Best Local Similarity 19.9%; Pred. No. 10;
Matches 69; Conservative 69; Mismatches 113; Indels 96; Gaps 16;

QY 4 IPAEQTVVVVKSATETADGALDLYNKYLDQVTPWKTETETIKELSRPK-QEYSQEAS-VL 61
DB 36 IFGEDEVLMN-----W--LNEVHDKLSKLSVQDYSTEGLWKQ 71
QY 62 VGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKQD----- 115
DB 72 QSELRVQEDILLRKQNVQDALLNGLELLKQTTGDEVLIIQDKLEAKARYKDIITKUSTD 131

; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 666
; LENGTH: 1318
; TYPE: PRP
; ORGANISM: Homo sapiens
; US-10-408-765A-666

Query Match      6.9%; Score 105; DB 14; Length 1174;
Best Local Similarity 19.9%; Pred. No. 13;
Matches 69; Conservative 69; Mismatches 113; Indels 96; Gaps 16;

QY 4 IPAEQTVVVVKSATETADGALDLYNKYLDQVTPWKTETETIKELSRPK-QEYSQEAS-VL 61
DB 761 IFGEDEVLMN-----W--LNEVHDKLSKLSVQDYSTEGLWKQ 796
QY 62 VGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKQD----- 115
DB 797 QSELRVQEDILLRKQNVQDALLNGLELLKQTTGDEVLIIQDKLEAKARYKDIITKUSTD 856
QY 116 LIRILDDGV---KLLNEAOKSLTSSQFNNSAGKLLALDLSQTLNDFSEKSSYFQSQVDR 172
DB 857 VAKTLEQALQARLHSTHEELCT---WLDKVEVELLSYETQVLK--GEEASQAMRPKE 911
```

RESULT 33
 US-10-437-963-135452
 ; Sequence 135452, Application US/10437963
 ; Publication NO. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 135452
 ; LENGTH: 2295
 ; TYPE: PRI
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT MRT4530_37127C.1.pcp
 ; US-10-437-963-135452

Matches 60; Conservative 63; Mismatches 120; Indels 76; Gaps 14;

7 EQTEVVVKSATETAD---GALD-----LYNKYLDQVVPKTFDETIKE-----LS 48
192 ETTVNDLRKSVESWTLSGLGAVENSITQATLSNIQQQTKELSNTRNISEMLDILSGMS 251
49 RFQOYSQOASVLVDGIKVLMDSDQKYFEATQTVYEWGCVVTOILSAYILLFDEYNEKK 108
252 NIQEISETKAFGGLHSEVVEDK-MDAT-----NQLITS---XPFDEFSELL 297
109 ASAQKDILIRLDDGVKLNKAEQKSLLT-----SSQSFNNASGKLLALDSOLT 156
298 KKSNTAEALVEVMKTVTEEFQKQNSLNKLKIQENFDQNLKSVKELNTWQOENKAMISLT 357
157 NDFSEKSYFOS---QVDRIRKEAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELAN 213
358 QOYKEMASNPSTSTTLISQVDDRTLVSEB---GKLKQLID-SLNQVIVEDQKFDVSN 413
214 RLKTVQNFPTS-----LSATVQANKDIDAACKLATEIAAIGIKTETETTRF 262
414 KLOETANISKEMKPFESTKILNDWVRKORNFVDGVQL-LIEKLDLKNIR----- 464
263 YVDYDDLMLSLKGAACKM 281
465 --DYSE---QFWKGTKEKM 478

RESULT 36
US-10-282-122A-51808
; Sequence 51808, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITEA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51808
; LENGTH: 809

TYPE: PRT
ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51808

Query Match 6.9%; Score 104; DB 15; Length 809;
Best Local Similarity 22.3%; Pred. No. 9.6;
Matches 67; Conservative 40; Mismatches 110; Indels 84; Gaps 14;

24 LDLYNKYLDQVVPKTFDETIKELSRFKQBYSEASVLVDGIKVLMDSDQKYFEATQTV 83
265 LVLYKRNQNVYDFSTYQNAIDLNKLAFSQO-----KI-----SNKY-----T 305
84 YEWCGV-----VTOLLGAYILLFDEYNEKASQAQKDLIRLDDGVKLNKAEQKSLTSS 138
306 YEWFSIPVNVQVQDLKSQVHYTDE-----BIDSLLR--DGGLKIYTTMTSMESNV 355
139 QSP--NNASGKLLALDSOLTNDSEKSYFOSQVDRIRKEAYAGAAAGIVAGPGLIISY 196
356 QNILDNNSTLK-----SYSYADKNGIIQPEAAATLFDVHTGEIKAIIVGG----- 399
197 SIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVQANKDIDAACKLATEIAAIGIKTE 256
400 -----RGOQPPGSSYNRADS--SNLYRSVGSISIKPLTYVAPAITKLATEDTIVNDSPLS 451
257 TETTRFY-----VDY-----DDLMLS-----LLKGAACKMIN-----TCNEYQORH 292
452 SDVAEKYGSNGVVPYHPNDGSGYPVNLKTALTKSINLVAIKLEDKLGISTGAAYAQKF 511
293 G 293
512 G 512

RESULT 37
US-10-369-493-3279
; Sequence 3279, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3279
; LENGTH: 1965
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3279

Query Match 6.9%; Score 104; DB 14; Length 1965;
Best Local Similarity 21.2%; Pred. No. 32;
Matches 63; Conservative 45; Mismatches 119; Indels 70; Gaps 11;

6 AEQTVVVVKSATETADGALDLYNKYLDQVVPKTFD-----ETIKELSRFKQBYSQ 56
57 AKYREELKQLEHAKEK-----NDWLNEUKTQNAEAKFKKKGARIKELQRLNEDANS 112
57 EASVLVDGIKVL---LMDSDQKYFEATQTVYEWG-----GVVTOLLSA---YILLFDE 103
113 TIESLTRSEQVLRKLIQEAQDKABETLTQVQLOEAAARTEEGFKQELSAKLVKLDQ 172
104 YNEKASQAQKDLIR---ILDDGVKLNKAEQKSLTSSQSFNNASGKLLALDSOLTNDFS 160
173 QSETHRNRLKEVELRLEQIKDDHANEIRIRRELEQKE-----DHAQT---E 217

```
QY 161 EKSSYFOSQVDRIRKE-----AYAGAAAGIVAGPGLIISYIAG 201
Db 218 QRAQELQNEVDRIKASSDLGRSPGTPQPKGDSFAATRAG---SPFGTPIUSIRGAG 274
QY 202 VIEGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAAKLKLATEIAAIGEEKTETE 258
Db 275 QRATDALEELYN---VKGQAGEKRCCKLQBELDDAVAMLEAKWPEIDELNAESE 327

RESULT 38
US-10-425-115-267777
; Sequence 267777, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 267777
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(568)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_175816C.1.pcp
US-10-425-115-267777

Query Match 6.8%; Score 103.5; DB 17; Length 568;
Best Local Similarity 21.4%; Pred. No. 6.5;
Matches 70; Conservative 51; Mismatches 115; Indels 91; Gaps 14;

QY 3 SIFAEQTVVVKSAIETADGALDLYNKYLQVVP--WKTPEIKELSRFKQYEQEASV 60
Db 9 SVFASGTAELSLTAQQA-----DVANKALQDNTRKXETHKTVTSALVK--TDTHLNEALS 62
QY 61 LVGDIKVLMDSQDKYFEATQTVYEWGVVTVLLSAYILLDFBYNEKKSAAQKDIILIRIL 120
Db 63 EISSLESGLQDAEKRF-----VTMQELRDYISVMCDFLNDKA-----FLIEEL 105
QY 121 DDGVKVLNEAQKSLTSSQSFNNSGKLLALDSQLTNDPSEKSSYFQSQVDRIKAYAG 180
Db 106 EENIQQLHEK-----RALAISERRAADLADESGVI-----EAAVS 140
QY 181 AAAGIVA-GPFGLIISYS-----IAAGVIEG--KLIPEL-----N 212
Db 141 AAVSILSKGSSSTCLSNASNAQAQAAAAGSSNLQPELDEFGRDINMQKMDLKRRED 200
QY 213 NRLKTVQNFPTSLSATVKQANKDIDAAKLKLATEIAAIGIKTETETTRFYVDYDMLMS 272
Db 201 RRQRKTQSETKRLASAAK--NKDIKKIEGELSTD-----ESDSESTAVVSRDE---- 247
QY 273 LLKGAAKKMTNCNEYQQRHGKKTLP 299
Db 248 FLKAAOHVFDIAKEEYSSSLRIVKDKPE 274

RESULT 39
US-10-408-765A-1635
; Sequence 1635, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
```

```
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1635
; LENGTH: 1583
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-1635

Query Match 6.8%; Score 103.5; DB 16; Length 1583;
Best Local Similarity 19.7%; Pred. No. 26;
Matches 69; Conservative 54; Mismatches 89; Indels 139; Gaps 15;

QY 7 EQTVVVKSAIETADGALDLYNKYLQVVPWKT--FDETIKELSRFKQYEQEASVILVGD 64
Db 531 EQTIQ-----YNSELRQKVNELTGGLEETLKE-----KDQNDQKLEKLMVQ 571
QY 65 IKVLLMDSQDKYFEATQTVYEWGVVTVLLSAYI--LLDFEYNEKKSAAQKDIILIRILD-- 121
Db 572 MKVL-----SEDK-----EVLSAEVKSLYEENN--KLSSSEKKQLSRDLVLF 610
QY 122 -----DGVKKL--NEAOKSLTSSQSFNNSGKLLALD 152
Db 611 LQKEDVILKEHTIQLEKKLQLMVERQDNKLLENQVKLFVKTLQY----GFLKEMG 666
QY 153 SOLTNDPSEKSSYFQSQVDRIKAYAGAAAGIVAGPGLIISYIAGVIEGKL----- 207
Db 667 SEVSEDSSEKD-----VNVLQAVGESLAKINEKC 697
QY 208 -----IPELNNRLKTVQNFPTSLSATVKQANKDIDAAKLKLATEIAAIG----- 251
Db 698 NLAQFQDEKYLEKEIKCLQESVYVQCBEKSLRLDYEQEKVLLRKELEEIQSEKEALQ 757
QY 252 ----BIKTETETTRFYVDYDMLMSLLKGAAKKMTNCNEYQQRHGKKTLP 298
Db 758 SDLLEMMANNEKTR--LENQNLIQV-----EEVSQTCSEIHNEKEKCF 801

RESULT 40
US-10-437-963-124349
; Sequence 124349, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 124349
; LENGTH: 1585
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1585)
; OTHER INFORMATION: unsure at all Xaa locations
```

; PRIOR APPLICATION NUMBER: 60/311266
 ; PRIOR FILING DATE: 2001-08-09
 ; NUMBER OF SEQ ID NOS: 211
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 36
 ; LENGTH: 1959
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-028-248A-36

Query Match 6.8%; Score 103.5; DB 14; Length 1959;
 Best Local Similarity 19.2%; Pred. No. 35;
 Matches 68; Conservative 49; Mismatches 119; Indels 119; Gaps 13;
 QY 7 EQTVVVKSAIE-----TADGALDLYNKYLDPVTPKTFDETIKELSRFKQEYS 55
 DB 1166 EQVNVILKTLLEAKTHEAQIQEMRQKHSQAV--BELAEQLQETKRVKANLEKAKOTLE 1223
 QY 56 QASVVLGVDIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKSASQKDI 115
 DB 1224 NERGELANEVVKVLLQGGD-----SEHKRKVEAQLQE 1256
 QY 116 LIRILDDGVKVLNEAOKSLITSSQSFNNASGKLALD---SOLTNDPSEKSSYFQSOVDR 172
 DB 1257 LQVFNQEGERVTELADKVTKLQVELDNTGLLSQSDSKSKLTDFSALESQLODQTEL 1316
 QY 173 IRKEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNRLKTVQ---NFF----- 222
 DB 1317 LOEEN-----RQKL--SLSTKLKQVEDEKNSFREOLEEE 1348
 QY 223 -----TSLSATV-----KQAKDIDAUKLATEIAAIG 251
 DB 1349 EAKHNLKQIATLHAQVADMKKQMEDSVGCLTAEVVKRLQKDLGLSQRHEEKVAAYD 1408
 QY 252 EIKTETETFRFYVDYDDMLMLSLKGAAKMINTCN-EYQORHGKTKLFEVDPVAS 305
 DB 1409 --KLEKTKRLQOELDLVLDL-----DHORASACNLEKKQKFPOLLAEKTTISA 1457

RESULT 42
 US-10-107-782-36
 ; Sequence 36, Application US/10107782
 ; Publication No. US20040018970A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boldog, Ferenc,
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steve,
 ; APPLICANT: Edinger, Shlomit,
 ; APPLICANT: Gangolli, Esha,
 ; APPLICANT: Kekuda, Ramesh,
 ; APPLICANT: Li, Li,
 ; APPLICANT: Liu, Xiaohong,
 ; APPLICANT: Malvankar, Uriel,
 ; APPLICANT: Miller, Charles,
 ; APPLICANT: Millet, Isabelle,
 ; APPLICANT: Patturajan, Meera,
 ; APPLICANT: Rothenberg, Mark,
 ; APPLICANT: Sciore, Paul,
 ; APPLICANT: Shenoy, Suresh,
 ; APPLICANT: Shimkets, Richard,
 ; APPLICANT: Smithson, Glenda,
 ; APPLICANT: Spytek, Kimberly,
 ; APPLICANT: Stone, David,
 ; APPLICANT: Taupier, Raymond, Jr.,
 ; APPLICANT: Tchernev, Velizar,
 ; APPLICANT: Vernet, Corine,
 ; APPLICANT: Zerhusen, Brian
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
 ; FILE REFERENCE: 21402-222CIP
 ; CURRENT APPLICATION NUMBER: US/10/107,782
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: 10/028,248

; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_27097C.1.pap
 ; US-10-437-963-124349

Query Match 6.8%; Score 103.5; DB 16; Length 1585;
 Best Local Similarity 22.1%; Pred. No. 26;
 Matches 61; Conservative 48; Mismatches 114; Indels 53; Gaps 10;
 QY 2 TSIFAEQ--TVEVKSATADGALDLYNKYLDPVTPKTFDETIKELSRFKQEYSQAS 59
 DB 608 TSLLAEKQHAALMSQLAETKQNEELQKLDVN---RTNDLQDSLKRPENVTTRDA 664
 QY 60 VLVG-----DIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKSASQK 113
 DB 665 LYLAEQHEDETQKSLSKSQERNWELLQKVDEAKRINKLENAQRL-----EKHATAKE 719
 QY 114 DILIRIL---DDGVKVLNEAOKSLITSSQSFNNASGKLALD---SOLTNDPSEKSSY-- 165
 DB 720 SLLKTKQSHDSTTKALVEAESNRNRELTQSFSDRKINLLEDSVNRLEERIAEKDLSLE 779
 QY 166 FQSOVDRIKRAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNRLKTVQNFSTSL 225
 DB 780 IERQENNAKDEVINA-----QNKIM-ELVNESQLOQDIRKHL 816
 QY 226 SATVQAKNDIDAUKLATEIAAIGLTKTETTR 261
 DB 817 EDNIKRLEED-----ATTREALLISEKQTHEATKR 846

RESULT 41
 US-10-028-248A-36
 ; Sequence 36, Application US/10028248A
 ; Publication No. US20030235882A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Vernet, Corine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Malvankar, Uriel
 ; APPLICANT: Shenoy, Suresh
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Gangolli, Esha
 ; APPLICANT: Miller, Charles
 ; APPLICANT: Boldog, Ferenc
 ; APPLICANT: Li, Li
 ; APPLICANT: Taupier Jr, Raymond J
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Smithson, Glenda
 ; APPLICANT: Zerhusen, Bryan
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Tchernev, Velizar
 ; APPLICANT: Si, Jingsheng
 ; APPLICANT: Edinger, Shlomit
 ; APPLICANT: Stone, David
 ; APPLICANT: Sciore, Paul
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Rothenberg, Mark
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS
 ; FILE REFERENCE: 21402-222
 ; CURRENT APPLICATION NUMBER: US/10/028,248A
 ; CURRENT FILING DATE: 2001-12-19
 ; PRIOR APPLICATION NUMBER: 60/256619
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: 60/262959
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: 60/272408
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: 60/285189
 ; PRIOR FILING DATE: 2001-04-20
 ; PRIOR APPLICATION NUMBER: 60/308039
 ; PRIOR FILING DATE: 2001-07-26

Publication No. US20030235882A1
GENERAL INFORMATION:
APPLICANT: Shinkets, Richard
APPLICANT: Patturajan, Meera
APPLICANT: Vernet, Corine
APPLICANT: Casman, Stacie
APPLICANT: Malyankar, Uriel
APPLICANT: Shenoy, Suresh
APPLICANT: Spytek, Kimberly
APPLICANT: Gangolli, Esha
APPLICANT: Miller, Charles
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Kekuda, Ramesh
APPLICANT: Smithson, Glennnda
APPLICANT: Zethusen, Bryan
APPLICANT: Liu, Xiaohong
APPLICANT: Colman, Steven
APPLICANT: Tchernev, Velizar
APPLICANT: Si, Jingsheng
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Sciore, Paul
APPLICANT: Millet, Isabelle
APPLICANT: Rothenberg, Mark
TITLE OF INVENTION: NO. US20030235882A1el Nucleic Acids and Polypeptides and Methods
TITLE OF INVENTION: Thereof
FILE REFERENCE: 21402-222
CURRENT APPLICATION NUMBER: US/10/028,248A
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/256619
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 60/262959
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/272408
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/285189
PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: 60/308039
PRIOR FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 60/311266
PRIOR FILING DATE: 2001-08-09
NUMBER OF SEQ ID NOS: 211
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 103
TYPE: PRT
ORGANISM: Homo sapiens
US-10-028-248A-103
Query Match 6.8%; Score 102.5; DB 14; Length 1961;
Best Local Similarity 19.8%; Pred. No. 42; Indels 87; Gaps 11;
Matches 60; Conservative 47; Mismatches 114; Indels 87; Gaps 11;
Qy 7 EOTVEVVKSAIE---TADGALDLYNKYLDQVPIPKTPTETIKE-----LSRFQOYVS 55
Db 1166 EQVNLKKTLEBEAKTHEAQIQEMRQKSOAV--BELAQEQTKRVKANLEKAKQTLE 1223
Qy 56 QEASVLVDGDKVLMDSQDKYFATQTVYEWCGVVTQLLSAYILLFDEYNEKKSAAQKDI 1115
Db 1224 NERGELANEVKVLLQGRD-----SEHKKKVEAQLQE 1256
Qy 116 LIRILDGVKKLNEAQKSLTSSQFNNSGKLIALD---SQTNPFSKSSYFQSQVDR 172
Db 1257 LQVFNESGVRVTEADKVKYQLVDELNVGTLSSQSDSKSKLTQPSALESQDQTOEL 1316
Qy 173 IRKEAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNNRLKTVQ----NFTSLSAT 228
Db 1317 LQEE-----RQKL--SLSTKLQVDEKNSFREQLLEE 1348
Qy 229 VKQAKDIDAAKKLATEIAAIGIKTETETTRFYVDYDDL-MLSLKGAAGKMCNTCNE 287

US-10-437-963-176713
Sequence 176713, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 176713
LENGTH: 957
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_74436C.1.pap
US-10-437-963-176713
Query Match 6.8%; Score 102.5; DB 16; Length 957;
Best Local Similarity 21.2%; Pred. No. 16; Indels 51; Gaps 11;
Matches 60; Conservative 53; Mismatches 119; Indels 51; Gaps 11;
Qy 7 EOTVEVVKSAIETADGALDLYNKYLDQVPIPKTPTETIKELSRFQOYVSQASVVLGDIK 66
Db 23 KEGITFNLSELSMQAEVDKISK-----VPLQDLSQIKIWARDVRELSYDIEDNV-DTF 76
Qy 67 VLMDSDQ--KYFEATQTVYEWCGVVTQI-----LSAYILLFDEYNEKK 108
Db 77 MLCVDDLEARKKHDFWLIDKTKYCSLSLKHKIHKIANDIKHDIIIPVKEVVERHRYNADD 136
Qy 109 ASAQDILIRILDGVKKLNEAQKSLTSSQFNNSGKLIALDLSQLTNDFSEKSSYFQS 168
Db 137 VDSK---LPTIIDPRILKLYDN---VTKPVGVDRKASGLIKLSMETDSSQK----- 183
Qy 169 QVDRIRKEAYAGAAAGIVAGP--FGLI--ISYIAAGVIEGKLIPELNNRLKTVQNTFTLS 226
Db 184 -LKMISVVGPGGLGKTTLAKEVFMRLRVQFSVACFSVGRK-PDKKKVLKSI----- 233
Qy 227 ATVKQAKDIDAAKKLATEIAAIGIKTETETTRFYVDYDDL 269
Db 234 --LIEVNVKQKMSDLAKSLSERHLIDREYLENRYLVLLDDI 274
RESULT 46
US-10-028-248A-103
Sequence 103, Application US/10028248A

Db 1349 EBEAKHNLEK--QIATLHAQVADMKKME-----DSVGCLETAEEVKRLKQDLEG 1397
QY 288 YQORHGKK 295
Db 1398 LSQRHEEK 1405

RESULT 47

US-10-107-782-103
; Sequence 103, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Perenc,
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Gangolli, Baha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li,
; APPLICANT: Liu, Xiaohong,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2001-12-19
; PRIOR FILING DATE: 2001-01-19
; PRIOR FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2001-04-20
; PRIOR FILING DATE: 2001-07-26
; PRIOR FILING DATE: 2001-07-26
; PRIOR FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: CuraSeglist version 0.1
; SEQ ID NO 103
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-103

Query Match 6.8%; Score 102.5; DB 15; Length 1961;
Best Local Similarity 19.5%; Pred. No. 42;
Matches 60; Conservative 47; Mismatches 114; Indels 87; Gaps 11;
QY 7 EQTVEVVKSAIE----TADGALDLYNKYLDQVVPWKTFDETIKE-----LSRFRKQEYS 55
Db 1166 EQEWNLKKTLEBEAKTHAQIQEMRKHSQAV--BELAQLEQTKRVKANLEKAKQTL 1223
QY 56 QEASVLVDIKVLLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASAKQDI 115

Db 1224 NERGELANEVKKVLLQGRD-----SEHKRKKVEAQLOE 1256
QY 116 LTRILDGDKVKKLNEAQKSLTSSQSFNNASGKLALD---SOLTDFSEKSSYFQSOVDR 172
Db 1257 LQVKFNEGERVRELADKVTQKQVELDQVNTGLSSQSDSKSKLTQFSALESQLODQTEL 1316
QY 173 IRKEAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQ---NFTSLSAT 228
Db 1317 LOEEN-----ROKL--SLSTKLQKQVEDEKNSFRELEEE 1348
QY 229 VKQANKDIDAAKLKLATEIAAIGEIKETETETTFYVDYDDL-MLSLLKGAAKKMWINTCNE 287
Db 1349 EBEAKHNLEK--QIATLHAQVADMKKME-----DSVGCLETAEEVKRLKQDLEG 1397
QY 288 YQORHGKK 295
Db 1398 LSQRHEEK 1405

RESULT 48

US-10-369-493-2102
; Sequence 2102, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkie, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2102
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2102

Query Match 6.7%; Score 102; DB 14; Length 451;
Best Local Similarity 21.6%; Pred. No. 6.4;
Matches 81; Conservative 45; Mismatches 131; Indels 118; Gaps 15;
QY 7 EQTVEVVKSA-----LETADGALDLYNKYLDQVVPWKTFDETIKELSRKQEY--- 54
Db 4 EENVKEAKNAFNLTQLSVEDRDDAL-----DKIVEELRIKKSEVLAAAN 47
QY 55 --SQEASVLVGD-----IKVLLMSQDKYFEATQTVYEW-----CGVVTQLLSAYI 98
Db 48 AEDMKAALKLAESGKLSNMVKELDLSSSDKYSWQGVLDVKSLLPDLGRVTVARS--- 104
QY 99 LFPDEYNEKKA-----AQXILIRILDGQVKKLN-----BAQSKLLTSSQ 139
Db 105 -LDDGLDLYKVCVGLLVIFEARPEVIINITSLAISKGNVAVLVKGGTSAKSAFALSN 163
QY 140 SFNNASGK-----LLALDSQLTDFSEKSSYFQSOVDRIKAEYAG 180
Db 164 VRSALGKSKVPOAVOLVQSREEVSQLKLDBYIDLVI PRGSTNLVHRHKDNTKIPVLG 223
QY 181 AAAGI-----VAGPFGI--IISYSIAAGVIEGKLIPE--LNNRLKTVQNFPTS 224
Db 224 HAAGLCMYVHEDADDELASKLVLDGKTDYPAACNAIETLLINEAVLSSHLPKIAETL 283
QY 225 LSATVQANKDIDAAKLKLATEIAAIGEIKETETTFYVDYDDLMLSLK-----GAACK 280
Db 284 AKVTLK-----CDPASLKVLDKMPKVSALVPEPSVDQDYNTEFSDILAIKTVPSLQSA 338
QY 281 MINTCNEYQORHGKK 295

LENGTH: 1038
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7776

Query Match 6.7%; Score 102; DB 14; Length 1038;
Best Local Similarity 20.8%; Pred. No. 20; Indels 66; Gaps 13;
Matches 64; Conservative 59; Mismatches 118; Indels 66; Gaps 13;

QY 7 BOTVEVVKSS---ALETADGALDLYNKYLDQVVPKTFDETIKELSRPKQB-YSQEASVLV 62
Db EETIQALKSERNELTAKVSELEDYMKHSE-----VEPDVVMKQNDQFQERIHELEAIDT 291
QY 63 GDIKVLMDSDQKYFEATQTVYEWGCVVTTQLLSAYILLFDEYNEKKAQKIDILRLD- 121
Db LHQTEATIQQSOSRENTE-----LQLOKLTSTELDKQOEMNRLKASKNENLEMDLSEK 344
QY 122 -DGVKCLNEAQSLLTSSQSFNNASGKLALDSQLTN--DFSEK-SSYFQSOVDRIK-KE 176
Db TDNLKELN---NKVLSQAEINILLETKLDLTNSQFENNTDGNKLMKNLESQNKVQTOE 401
QY 177 AYAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNPFTSLSATVQANKDI 236
Db 402 AF-----IDELHHEQKTIIDNEY-----KAKIKDL 425
QY 237 DAAKUKLATEIAAIGETETETETTFYVYDDML--SLKGAAKKMMINTCNEYQOR--- 291
Db 426 EYENAQLSEISIRAKNSQYDPEAQHYEIDQLQENALQKDNVKNYLNFKELKDKVE 485
QY 292 HGKKTIF 298
Db 486 HAHQIAF 492

Search completed: January 5, 2005, 11:08:47
Job time : 66.7056 secs

Db 339 HINT-----HGSK 346

RESULT 49
US-10-369-493-3210
Sequence 3210, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 3210
LENGTH: 1015
TYPE: PRT
ORGANISM: Neurospora crassa
US-10-369-493-3210

Query Match 6.7%; Score 102; DB 14; Length 1015;
Best Local Similarity 20.5%; Pred. No. 19;
Matches 65; Conservative 57; Mismatches 123; Indels 72; Gaps 12;

QY 18 ETADGALDLYNKYL-DQVVPKTFDETIKELSRPKQESQASVLVGDIVL-----LM 70
Db 377 EDPSSALADQKSLQDQDRSEKKNALNQKITNLKNSIEBANVALDKMLQVVERSLV 436
QY 71 DSQDK---YFEATQTVYEWGCVVTTQLLSAYILLFD-----EYNEKKAQAKDI 115
Db 437 DLQQRQEGDKSLNQVLDLQKASKASAGADLKDIOQLRLNKSLEDFQORVQAEKV 496
QY 116 LRIILDDGVKCLNEAQSLLTSSQSFNNASGKLALDSQLTNDFSEKSSYFQSOVDRIK 175
Db 497 LQQLSQTARLEKVEITL-----KNTASQMDLETQ-RNEWSKAKKVLREISRLKK 548
QY 176 EYAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLK-----TVQN 220
Db 549 EV-----DLARKNSAAPTIIVEIKPEPNIAESITRVEQLMKEQKPLPLTARD 595
QY 221 FFTLSATVK--QANKIDIAAKLATEIAAIGETETETTFYVYDDMLSLKGA 278
Db 596 FVQGLVAKLKDAQGHKDLDAEDVIRECIGLLNRLKEP-----VFDEKDKEMLQTLQ--- 648
QY 279 KMINTCNEYQQRHGKK 295
Db 649 ----NELNE--ALRKGKE 660

RESULT 50
US-10-032-585-7776
Sequence 7776, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jiang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7776

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:43:53 ; Search time 55.6826 Seconds
(without alignments)
3151.602 Million cell updates/sec

Title: US-09-993-292B-2
Perfect score: 1515
Sequence: 1 MTSIFAEQTEVVVKSIAETA.....NEYQQRHGKTLFEVDPVAS 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : UniProt 02:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1498	98.9	302	1 HLYE_SALTI	Q8727 salmonella
2	1461	96.4	302	1 HLYE_SALPA	Q9316 salmonella
3	1397	92.2	302	1 HLYE_ECOLI	P7735 escherichia
4	1382	91.2	302	1 HLYE_ECO57	Q9reb3 escherichia
5	1110	73.3	300	2 Q9X2S8	Q9x288 escherichia
6	519	34.3	113	1 HLEL_SHIFL	Q9rct3 shigella fl
7	395	26.1	93	1 HLEL_ECOL6	Q8fi27 escherichia
8	130.5	8.6	895	2 Q9LIW7	Q9liw7 oryza sativ
9	128.5	8.5	4007	2 Q7SHZ4	Q7shz4 neurospora
10	126	8.3	1521	2 Q6G015	Q6g015 bartonella
11	123.5	8.2	713	2 Q6MT03	Q6mt03 mycoplasma
12	123.5	8.2	713	2 CAE77235	Caet77235 mycoplasma
13	122.5	8.1	693	2 Q7M918	Q7m918 wolpinella s
14	122	8.1	495	2 P71497	P71497 mycoplasma
15	121	8.0	2723	2 Q7RQB6	Q7rqb6 plasmodium
16	119	7.9	541	2 Q74DE2	Q74de2 geobacter s
17	119	7.9	541	2 AAR34750	Aar34750 geobacter
18	118	7.8	1496	2 Q9SZK7	Q9szk7 arabidopsis
19	118	7.8	1877	2 Q6BFD6	Q6bfd6 parametium
20	117.5	7.8	577	2 Q7NSU5	Q7nsu5 photorhabd
21	117.5	7.8	1224	2 Q7QQ04	Q7qq04 giardia lam
22	116	7.7	1023	1 HLYI_ECOL1	P09983 escherichia
23	116	7.7	2033	2 Q7XEH4	Q7xeh4 oryza sativ
24	115.5	7.6	622	1 SR68_CAEEL	Q20822 caenorhabdi
25	115	7.6	587	2 Q97WH8	Q97wh8 sulfobolus
26	114.5	7.6	478	2 Q64067	Q64067 bacterioph
27	114.5	7.6	478	2 Q31954	Q31954 bacillus su
28	114.5	7.6	588	2 Q9VGA8	Q9vga8 drosophila
29	114.5	7.6	652	2 Q7CMF0	Q7cmf0 bacillus an
30	114.5	7.6	652	2 Q9X360	Q9x360 bacillus an
31	114.5	7.6	652	2 AAT28865	Aat28865 bacillus

ALIGNMENTS

RESULT 1

* HLYE_SALTI STANDARD; PRT; 302 AA.

```

32 114.5 7.6 779 2 Q7RM79
33 114.5 7.6 1885 2 Q869B8
34 114 7.5 1720 2 Q9HGQ7
35 114 7.5 1189 2 Q8CFI0
36 114 7.5 1611 2 Q7RD43
37 113.5 7.5 538 2 Q9Z429
38 113.5 7.5 727 2 Q8Z0N2
39 113.5 7.5 1127 2 Q9VVT6
40 113.5 7.5 1363 2 Q874Y4
41 113.5 7.5 1489 2 Q7RJ32
42 113.5 7.5 1906 2 Q6C359
43 113.5 7.5 3102 2 Q45614
44 113 7.5 465 2 Q73AJ3
45 113 7.5 465 2 AAS40708
46 113 7.5 956 1 YEF3_YEAST
47 113 7.5 1005 1 RAS0_METJA
48 113 7.5 1628 1 NAGH_CLOPE
49 112.5 7.4 927 2 Q71X69
50 112.5 7.4 927 2 AAT05096

```

AC Q8Z727; Q934C4;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA)
GN Names-hlyE; Synonyms=clyA, sheA; OrderedlocusNames=STY1498, t1477;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCSI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2la, and SMI S2369/96;
RX MEDLINE=22215712; PubMed=12228306;
RA Oscarsson J., Westermarck M., Loeffdahl S., Olsen B., Palmgren H., Mizunoe Y., Wai S.N., Uhlin B.E.;
RT "Characterization of a pore-forming cytotoxin expressed by Salmonella enterica serovar typhi and paratyphi A.";
RL Infect. Immun. 70:5759-5769(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -I- FUNCTION: Toxin, which has some hemolytic activity towards mammalian cells. Acts by forming a pore-like structure upon contact with mammalian cells (By similarity).

RESULT 3

HYLX ECOLI STANDARD; PRT; 302 AA.
 AC P77335; Q47276; OBVU70; 09R3G4;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 29-MAR-2004 (Rel. 43, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Hemolysin E, chromosomal (Hemolysis-inducing protein) (silent
 DE hemolysin sheA) (Cytotoxin clyA) (Latent pore-forming 34 kDa
 DE hemolysin)
 GN NameshlvE: Synonyms=clyA, sheA, hpr; OrderedLocusNames=bll182;
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / XLI-BLUE;
 RA McNamara P.J., Iandolo J.J., Uhlich G.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;
 RT "The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted
 RT haemolysin";
 RL Mol. Microbiol. 25:107-115(1997).
 RN [3]
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-12, SUBCELLULAR LOCATION, AND
 RP INDUCTION.
 RC STRAIN=K12;
 RX MEDLINE=99157562; PubMed=10027972;
 RA Ludwig A., Bauer S., Benz R., Bergmann B., Goebel W.;
 RT "Analysis of the SlyA-controlled expression, subcellular localization
 RT and pore-forming activity of a 34 kDa haemolysin (ClyA) from
 RT Escherichia coli K-12";
 RL Mol. Microbiol. 31:557-567(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3030-2;
 RA Xing J., Fernandez S.V., Kapur V., Barletta R.G., Moxley R.A.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1245-1247(1997).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:137-155(1996).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH9802;
 RA Chang G.-N., Ho K.-C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [8]

RP SEQUENCE OF 1-295 FROM N.A.
 RC STRAIN=K12 / XLI-BLUE;
 RA King C.H., Shinnick T.M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE OF 1-155 FROM N.A.
 RC STRAIN=K12 / AB1157;
 RA Woodgate R.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP SEQUENCE OF 1-17, AND MUTAGENESIS OF 87-GLY--VAL-89; 142-ASN-ALA-143;
 RP 182-ALA-GLY-183; 186-ALA-GLY-187; ASP-267 AND 292-GLY-LYS-293.
 RX MEDLINE=99316011; PubMed=10383763;
 RA Oscarson J., Mizunoe Y., Li L., Lai X.-H., Wieslander A., Uhlin B.E.;
 RT "Molecular analysis of the cytolytic protein ClyA (SheA) from
 RT Escherichia coli";
 RL Mol. Microbiol. 32:1226-1238(1999).
 RN [11]
 RP PARTIAL SEQUENCE, FUNCTION, SUBCELLULAR LOCATION, OLIGOMERIZATION, AND
 RP DISULFIDE BOND FORMATION.
 RX MEDLINE=22894281; PubMed=14532000;
 RA Wai S.N., Lindmark B., Soederblom T., Takade A., Westermarck M.,
 RA Oscarson J., Jass J., Richter-Dahlfors A., Mizunoe Y., Uhlin B.E.;
 RT "Vesicle-mediated export and assembly of pore-forming oligomers of the
 RT enterobacterial clyA cytotoxin";
 RL Cell 115:25-35(2003).
 RN [12]
 RP MASS SPECTROMETRY, DISULFIDE BOND, AND MUTAGENESIS OF TYR-96; ASN-156;
 RP TYR-164 AND ARG-260.
 RX MEDLINE=20576278; PubMed=11006277; DOI=10.1074/jbc.M005420200;
 RA Atkins A., Wyborn N.R., Wallace A.J., Stillman T.J., Black L.K.,
 RA Fielding A.B., Hisakado M., Artymluk P.J., Green J.;
 RT "Structure-function relationships of a novel bacterial toxin,
 RT hemolysin E. The role of alpha G";
 RL J. Biol. Chem. 275:41150-41155(2000).
 RN [13]
 RP INDUCTION.
 RX MEDLINE=20507803; PubMed=11053378;
 RA Westermarck M., Oscarson J., Mizunoe Y., Urbonaviciene J., Uhlin B.E.;
 RT "Silencing and activation of ClyA cytotoxin expression in Escherichia
 RT coli";
 RL J. Bacteriol. 182:6347-6357(2000).
 RN [14]
 RP INDUCTION.
 RX MEDLINE=22053220; PubMed=12057949;
 RA Spory A., Bosserhoff A., von Rhein C., Goebel W., Ludwig A.;
 RT "Differential regulation of multiple proteins of Escherichia coli and
 RT Salmonella enterica serovar Typhimurium by the transcriptional
 RT regulator SlyA";
 RL J. Bacteriol. 184:3549-3559(2002).
 RN [15]
 RP MUTANT PMWK16 DEL.
 RX MEDLINE=22829856; PubMed=12949101;
 RA Wai S.N., Westermarck M., Oscarson J., Jass J., Maier E., Benz R.,
 RA Uhlin B.E.;
 RT "Characterization of dominantly negative mutant ClyA cytotoxin
 RT proteins in Escherichia coli";
 RL J. Bacteriol. 185:5491-5499(2003).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20123445; PubMed=10660049;
 RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,
 RA Green J., Artymluk P.J.;
 RT "E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of
 RT the toxin and observation of membrane pores by electron microscopy";
 RL Cell 100:265-276(2000).
 CC -1- FUNCTION: Toxin, which has some hemolytic activity towards
 CC mammalian cells. Acts by forming a pore-like structure upon
 CC contact with mammalian cells.
 CC -1- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a
 CC monomer, while in outer membrane vesicles, it oligomerizes to form
 CC a pore structure that is active. Probably forms an octamer.
 CC -1- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer


```

CC -I- SIMILARITY: Belongs to the hemolysin E family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ238954; CAB64962.1; ALT INIT.
CC EMBL: AE055335; AAG56033.1; ALT INIT.
CC EMBL: AP002555; BAB35100.1; ALT_INIT.
CC PIR: E85696; E85696.
CC PIR: E90838; E90838.
CC HSSP: P77335; IQOY.
CC InterPro: IPR010356; HlyE.
CC Pfam: PF06109; HlyE; 1.
CC Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
FT INIT_MET 0 By similarity.
FT TRANSHEM 182 Potential.
FT DISULFID 86 In monomeric form (By similarity).
SQ SEQUENCE 302 AA; 33585 MW; F261B29B1B5FC87 CRC64;

Query Match 91.2%; Score 1382; DB 1; Length 302;
Best Local Similarity 90.1%; Pred. No. 1.8e-84;
Matches 272; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 2 TSIFAEQTVWVKSIAETADGALDLYNKYLDQVTPWKTFTDIKELSRFKQEYSQASVL 61
DB 1 TEIVADKTVVKNALETADGALDLYNKYLDQVTPWKTFTDIKELSRFKQEYSQASVL 60

QY 62 VGIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRILD 121
DB 61 VGNIKTLMDSDQKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKASAKDILIKVLD 120

QY 122 DGVKLENAOKSLTSSQSFNNASGKLALDSQLTNDSEKSSYFQSOVDRIKEAYAGA 181
DB 121 DGIITLENAOKSLTSSQSFNNASGKLALDSQLTNDSEKSSYFQSOVDRIKEAYAGA 180

QY 182 AAGVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTQNFFTLSATVQKAKDIDAACL 241
DB 181 AAGVAGPFGLLIISYSIAAGVIEGKLIPELKNKLSQSFVFTLSNTVQKAKDIDAACL 240

QY 242 KLATEIAAIGIKTETTRTFYVDYDDLMLSLKGAARKMINTCNEYQQRHGKTLFEVP 301
DB 241 KLATEIAAIGIKTETTRTFYVDYDDLMLSLKGAARKMINTCNEYQQRHGKTLFEVP 300

QY 302 DV 303
DB 301 EV 302

RESULT 5
QX288 PRELIMINARY; PRT; 300 AA.
AC QX288;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hemolysin.
GN Name=hlyE;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=964660;
RX MEDLINE=99242013; PubMed=10227474;
RA Reingold J., Starr N., Maurer J., Lee M.D.;
RT "Identification of a new Escherichia coli She haemolysin homolog in
RT avian E. coli.";

```

```

RL Vet. Microbiol. 66:125-134(1999).
DR EMBL: AF052225; RAD28079.1; -.
DR HSSP: P77335; IQOY.
DR InterPro: IPR010356; HlyE.
DR Pfam: PF06109; HlyE; 1.
SQ SEQUENCE 300 AA; 33555 MW; 6580B66C44A7B4BC CRC64;

Query Match 73.3%; Score 1110; DB 2; Length 300;
Best Local Similarity 74.1%; Pred. No. 2.7e-66;
Matches 215; Conservative 38; Mismatches 37; Indels 0; Gaps 0;

QY 6 AEQTVWVKSIAETADGALDLYNKYLDQVTPWKTFTDIKELSRFKQEYSQASVLVGI 65
DB 4 ADQTVETVTAIDTADKALDLYNKYLDQVTPWKTFTDIKELSRFKQEYSQASVLVGEI 63

QY 66 KVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRILDGVK 125
DB 64 KSLMNSQDRYFEATQTVYEWCGVVTQTLTAYLSLFENEYDEKKASAKDILIKVLDGII 123

QY 126 KLENAOKSLTSSQSFNNASGKLALDSQLTNDSEKSSYFQSOVDRIKEAYAGAAGI 185
DB 124 KLEKAOQSLHASSQSFNNASGKLALDSQLTNDSEKSDYFQGVQDKIRKEAYAGAAGV 183

QY 186 VAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTQNFFTLSATVQKAKDIDAACLKAT 245
DB 184 VGRPFGLLIISYSIAAGVIEGKLIPELKEKLSVKDFESLAATVKSANTDIDRAKSLKD 243

QY 246 EIAAIGIKTETTRTFYVDYDDLMLSLKGAARKMINTCNEYQQRHGK 295
DB 244 EISVIGDLKTETTRTFYVDYDDLMLKQLQDSATKLILSCNEYQQRHGK 293

RESULT 6
HLEL SHIFL
ID HLEL SHIFL STANDARD; PRT; 113 AA.
AC Q9RGT3; Q9RMG1;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin E-like protein.
GN OrderedLocusNames=SF1171, S1259;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12022 / Serotype 2b;
RX MEDLINE=20123445; PubMed=10660049;
RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,
RA Green J., Artymuk P.J.;
RT "E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of
RT the toxin and observation of membrane pores by electron microscopy.";
RL Cell 100:265-276(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CECT 585 / Serotype 2a;
RX MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;
RA del Castillo F.J., Moreno F., del Castillo I.;
RT "Characterization of the genes encoding the SheA haemolysin in
RT Escherichia coli O157:H7 and Shigella flexneri 2a.";
RL Res. Microbiol. 151:229-230(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";

```

```

RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RT Infect. Immun. 71:2775-2786 (2003).
CC -1- SIMILARITY: Belongs to the hemolysin E family.
CC -1- CAUTION: Although it is strongly related to the hemolysin E toxin
CC from E.coli K-12 strain, it lacks all the C-terminal part of the
CC protein, due to a deletion that creates a frameshift, and it is
CC therefore not functional. May be a pseudogene.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF200955; AAF13995.1; -
DR EMBL; AJ238955; CAB65415.1; ALT INIT.
DR EMBL; AE015143; -; NOT_ANNOTATED_CDS.
DR EMBL; AE016982; -; NOT_ANNOTATED_CDS.
DR HSSP; P77335; 10QY.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
DR SEQUENCE 113 AA; 12879 MW; C328908D14C5C4EB CRC64;
SQ
Query Match 34.3%; Score 519; DB 1; Length 113;
Best Local Similarity 91.0%; Pred. No. 2.9e-27;
Matches 101; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLDQVTPFTETIKELSRFKQYSQASV 60
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPFTETIKELSRFKQYSQASV 60
QY 61 LVGDIKVLMDSDQKFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
DB 61 LVGDIKVLMDSDQKFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
RESULT 7
HLEL_ECOL6 STANDARD; PRT; 93 AA.
AC QBF127;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin E-like protein.
GN OrderedLocusNames=cl630;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;
RX MEDLINE=22382234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RA "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
CC -1- SIMILARITY: Belongs to the hemolysin E family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF200955; AAF13995.1; -
DR EMBL; AJ238955; CAB65415.1; ALT INIT.
DR EMBL; AE015143; -; NOT_ANNOTATED_CDS.
DR EMBL; AE016982; -; NOT_ANNOTATED_CDS.
DR HSSP; P77335; 10QY.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
DR SEQUENCE 113 AA; 12879 MW; C328908D14C5C4EB CRC64;
SQ
Query Match 34.3%; Score 519; DB 1; Length 113;
Best Local Similarity 91.0%; Pred. No. 2.9e-27;
Matches 101; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLDQVTPFTETIKELSRFKQYSQASV 60
DB 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPFTETIKELSRFKQYSQASV 60
QY 61 LVGDIKVLMDSDQKFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
DB 61 LVGDIKVLMDSDQKFEATQVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
RESULT 7
HLEL_ECOL6 STANDARD; PRT; 93 AA.
AC QBF127;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin E-like protein.
GN OrderedLocusNames=cl630;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;
RX MEDLINE=22382234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RA "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
CC -1- SIMILARITY: Belongs to the hemolysin E family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF200955; AAF13995.1; -
DR HSSP; P77335; 10QY.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
DR SEQUENCE 93 AA; 10590 MW; CE1625028DC36DE6 CRC64;
SQ
Query Match 26.1%; Score 395; DB 1; Length 93;
Best Local Similarity 82.8%; Pred. No. 4.6e-19;
Matches 77; Conservative 9; Mismatches 7; Indels 0; Gaps 0;
QY 211 LNNRLKTVONFTSLSATVKQAKNDIDAAXIKLATEIAAIGCEIKETETTRFYVDYDLM 270
DB 1 MNKIKLSALSPFTLSNTVTKQAKNDIDAAXIKLATEIAAIGCEIKETETTRFYVDYDLM 60
QY 271 LSLKGAAKMINTCNEYQQRHGKTKLFEVPDV 303
DB 61 LSLKGAAKMINTCNEYQQRHGKTKLFEVPDV 93
RESULT 8
Q9LIW7 PRELIMINARY; PRT; 895 AA.
ID Q9LIW7;
AC Q9LIW7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Similar to an Arabidopsis thaliana chromosome BAC genomic
DE sequence.
DE Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC Hsing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001111; BAA90502.1; -
DR Gramene; Q9LIW7; -
DR SEQUENCE 895 AA; 100520 MW; BAA7BFDF70FPA3 CRC64;
SQ
Query Match 8.6%; Score 130.5; DB 2; Length 895;
Best Local Similarity 22.1%; Pred. No. 3.3;
Matches 76; Conservative 66; Mismatches 123; Indels 79; Gaps 17;
QY 8 QTVEVVKSAIETADGALDLYNKYLDQVTPFTETIKELSRFKQYSQASV 54
DB 274 EKVEILSVVRLKGLDSTAESEESKRETELV-KNLESEVSVLKGKLEARIIEER 331
QY 55 SQEASVLVGDIKVLMDSDQKFEATQVYEWCGVVTQLLSAYILLFDEYNEKKAS 110
DB 332 LAETKELIEELKSEVADAKKAESARQLFEWKHAGLLEMELEA-VTUSDKPKGSSLAS 390
QY 111 AQKDILRIIDDGVKKLNEAKQSLITSSQSFNNASGKLALD---SOLTNDPSEKSYF- 166
DB 391 T-----TEELGKIQSALQDRESEIEVLKGTITALEIEVALLADVNESNEQFD 438
QY 167 -----SQSDVRIRKEAYAGAAAGIVAGPFGLIISYSIAAGV-IEG-----KLIPE 210

```

```

Db 439 ASQEVFGLQTTIDVLRNKLAEAEAESEA-----LNNEKAANVKIEGLTEENVKLISE 492
Qy 211 LN-----NRLKTVQNFFTSLSTVQKANDIDAACKLATE-----IAAIGETKET 257
Db 493 LNETRDREEKRAVEDLTAAALS-----BESDKAKEAHERYLSKEDDHEHALAQIGDLKWL 549
Qy 258 ETRR-----FVVDYDMLSLKGAACKMINTCNEYQQR-HGKKT 296
Db 550 KSTKSEYVMDLEANYDITCLRNKVDKLEAEVKNYEECSKET 593

RESULT 9
Q7SHZ4
ID Q7SHZ4 PRELIMINARY; PRT; 4007 AA.
AC Q7SHZ4;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU00658.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins J., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Iankiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysellis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krysstofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0-0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000001; EAA36562.1; -.
DR InterPro; IPR009638; Pez1
DR InterPro; IPR003900; KID_repeat.
DR Pfam; PF06818; Pez1; 1.
DR Pfam; PF02524; KID; 5.
DR PROSITE; PS01039; SBP_BACTERIAL_3; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 4007 AA; 453246 MW; 67CDPF6EPF1463612 CRC64;

Query Match 8.5%; Score 128.5; DB 2; Length 4007;
Best Local Similarity 21.3%; Pred. No. 27;
Matches 75; Conservative 65; Mismatches 141; Indels 71; Gaps 14;

Qy 8 QTVVVKSAIETADGAL-----DLNKKYLDVIVPKTFDETIK-----ELSRKQE 53
Db 1156 QTEILIRKQHQSRGVELESRIATIKYKKLDLSELRNNTSQDAIKLKQHENELANPAK 1215
Qy 54 YSQEASLVGDIKVLMLDSQDKYPEATQTVYEWGCVVTVLLSAVI-----LLFDEYNKKA 109
Db 1216 YEQKKQLAVQHKTMEESLDRHYHEKELATQYQERV-QALSALADKKTALAEYKE-QL 1273
Qy 110 SAQKDILIRLDDGKVKLENAQKSL-----LTSSQSFNNAKGLALDSQ- 154
Db 1274 SASKAQLDKLKADHGKVKVDELQAKLKSEVAKVTADYEGNLSRLTRKHQGVNVLKHHQD 1333
Qy 155 -----LTWDFSEKSSYFQSQVDRIKKEAYAGAAAGVAGPGLIISYIAAGVIEGKL--- 207

```

```

Db 1334 EIKKLTAGHNEKIRNLEHRINDUKAELEKQDRA-----EFDKKKALDEGEVATL 1381
Qy 208 ---IPELNNRLKTVQNFFTSL-----SATVKQANKDIDAACKLATEIAAIGETKETET- 259
Db 1382 QGKVDKSSKSLKSEAEFNEKXLAELKADKADVADKSNLQDKLEELSDLKGGQCKTR 1441
Qy 260 -TRFYVDYDMLSLKLG-----AAKMIN-TCNEYQQRHG--KKTILFEVDP 302
Db 1442 IEDFNQINEKMAQLLKAQNELKASQASLNTTTTTEYDAKTAQLEKSLKEKD 1493

RESULT 10
Q6G015
ID Q6G015 PRELIMINARY; PRT; 1521 AA.
AC Q6G015;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=BQ05330;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RA PubMed=15210978;
RA Alenmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
RA Canbaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,
RA La Scolia B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
DR EMBL; BX897700; CAP26028.1; -.
DR InterPro; IPR009074; Apolipo_A_E_C3.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR010989; t-snare.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1521 AA; 169742 MW; 12CC93BCB2C97920 CRC64;

Query Match 8.3%; Score 126; DB 2; Length 1521;
Best Local Similarity 18.8%; Pred. No. 13;
Matches 70; Conservative 68; Mismatches 148; Indels 90; Gaps 8;

Qy 3 SIPAEQTVVVKSAIETADGALDLYNK-----YLDQVLPWK-----TFDE 42
Db 927 SVLSEQTAQTVESFTTASHNAQTLINETHTSATAIEEVLNERCNVLHSHMKNLGY 986
Qy 43 TIKELSRFKQYEQSEASVLVGDIKVLMLDSQDKYFEATQTVYEWGCVVTVLLSAIYLLFD 102
Db 987 QLSDSVNHLEESKQTATQISGHVEKLTETNYLNQAAQNTTESISHLTQHISEQLSLST 1046
Qy 103 EYNEKASAKQDILIRL-----DDGVKLNKAEQKSLTSSQSFNNASGK 147
Db 1047 QDAEQRIYAQNESLVNSLTQTNSETLTQVTAMKEDLVNNISILKQLNQSIYSPHNSNI 1106
Qy 148 LLALDSQLTNDFFSEKSSYFQSQVDRIKKEAYAGAAAGVAGPGLIISYIAAGVIE--- 204
Db 1107 LLSTVQNDQGFSETANNFRKTNQ-----AAEHLASNAQALNNNVLEVQLSQNIF 1158
Qy 205 -----GKLIPELNNRLKTVQNFFTSLSTVQKANDIDAACKLATEIAAIGEI- 253
Db 1159 EKIGHTITSTFGEHAKTSETIHILEKSENSLTSTLEBKHKTLTALSALVSKSNEINKLI 1218
Qy 254 -----KTETET-----TRFYVDYDMLSL--LKGRAKK 280
Db 1219 EYVENVLSLAFERTDTNTRNSTHSLQOQLNQLINEASTRFGAAEDIRRSADERSLSK 1278
Qy 281 MINTCYOQRHGKKT 296
Db 1279 INNDINESVQNLPEKT 1294

```

```

GN CLPB OR MSC_0613.
OS Mycoplasma mycoides (subsp. mycoides SC).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=44101;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PGL;
RY PubMed=14762060;
RA Westberg J., Persson A., Holmberg A., Goessmann A., Lundberg J.,
RA Johansson K.-E., Pettersson B., Uhlen M.;
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type
RT strain PGLT, the causative agent of contagious bovine pleuropneumonia
RT (CBPP).";
RL Genome Res. 14:221-227(2004).
RM EMBL; BX842644; CAE77235.1; -.
RW Protease.
RQ SEQUENCE 713 AA; 80782 MW; 1B5D204A9E29BE50 CRC64;

Query Match      8.2%; Score 123.5; DB 2; Length 713;
Best Local Similarity 23.3%; Pred. No. 7.5; Indels 95; Gaps 17;
Matches 79; Conservative 47; Mismatches 118;

QY 2 TSIPAEQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETI-----KELSRF 50
Db 396 TGLIVDRLLISSEKEKLNLELLKKYKVGQDAI--KAVTSAIMRSRSGIKNPKPIGSF 453
QY 51 -----KQEYSQF--ASVLVGDIKVLLMDSQDKYFEATQTV-----YEWCGV 89
Db 454 LFLGPTGVGKTEVARSLADILFNPKMIRLDMSEYMEKHSVAKLIGAPPGVGYEGGR 513
QY 90 VTQLL--SAY--ILLFDEYNEKKAQKIDILIRLLDG-----YKKLNE 129
Db 514 LTEAVRNPSYIVLFDEI--EKAHTDVFNILLQILDDGRLTDSLKGTIDFKNTIIVMTSNI 572
QY 130 AQKSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSQVDRIKKEAYAGAAAGIVAGP 189
Db 573 ASQYLLTSDE-----LVQVDDQ-----KIQELNKVFPFELNIDNIVY-- 612
QY 190 FGLIISYSIAAGVIEGKLIPELNNRLKTVQNFTSL--ATVKQAKDID-----AAKLK- 242
Db 613 FNALSQVTI--GEIVDKVLEELSTRLOEQNYFINFSEARNKIINEGYDRFLFGARPIKR 670
QY 243 -----LATEIAAIGEIKTETETTFYVYDDMLSL 273
Db 671 YIEKNIETLIAHYIISGEV---VENTRYLIDVKNNQFTL 706

RESULT 13
QY 190 FGLIISYSIAAGVIEGKLIPELNNRLKTVQNFTSL--ATVKQAKDID-----AAKLK- 242
Db 613 FNALSQVTI--GEIVDKVLEELSTRLOEQNYFINFSEARNKIINEGYDRFLFGARPIKR 670
QY 243 -----LATEIAAIGEIKTETETTFYVYDDMLSL 273
Db 671 YIEKNIETLIAHYIISGEV---VENTRYLIDVKNNQFTL 706

ID Q7M918 PRELIMINARY; PRT; 693 AA.
AC Q7M918;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN OrderedLocusNames=WS0888;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSMZ 1740;
RM MEDLINE=22882897; PubMed=14500908; Simon J., Lanz C., Klimmek O., B.,
RA Baar C., Eppinger M., Radatz G., Simon J., Lanz C., Klimmek O., B.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
RW EMBL; BX571659; CAB09993.1; -.
RQ Complete proteome.
KW SEQUENCE 693 AA; 77641 MW; CB57A1F23CFACC2 CRC64;
SQ

```

```
Query Match      8.1%; Score 122.5; DB 2; Length 693;
Best Local Similarity 21.7%; Pred. No. 8.4; Mismatches 115; Indels 77; Gaps 10;
Matches 67; Conservative 50;

Qy 11 EVKSAIETADGALDLYNKILQDVIPKTFDETIKLSRPFQYSQBSASVLGVGDKVLLM 70
Db 316 BEVTSSTKEAQESITQSRWLQRIQ-----SDISLAQQHTEHLGLSTFEKALV 367
Qy 71 DSQKYFEATQTVYEWGCVVTTLLSAYILLPDEYNEKKAQKDILRIIDDGVKKLNEA 130
Db 368 SLRQLGEGANEVLISLQKSSQESSAFLL---EQEKMAKESTIVLPAKLKERFERLSEA 423
Qy 131 QKSLLTSS-----QSFNASGKLLALDSOLTNDSEKSSYFQ 167
Db 424 ---FLTQNKGALEGMALDFRGFWEEYATRWRESELTOGTLLTNQQRASFAELSGVL 480
Qy 168 SQVDRIRKEAYAGAAAGVAGPF--GLIISYSIAAGVIEG--KLIPELN-----NRLKTQV 219
Db 481 AONSAMGEQIKAG-FEGMTQGALSGLMLAKALEGGIDGVKKSVMNQTLAKETLQTLQ 539
Qy 220 NFFTSLSATVKQANKIDAAKLIKLAIEAIGETITETTRFFVYDDLMLSLLKGAOK 279
Db 540 NHSQS-----QLGLLGESTTAIQR-----LLELEGAOK 569
Qy 280 KMINTCNEY 288
Db 570 SLKRLALEY 578

RESULT 14
P71497 PRELIMINARY; PRT; 495 AA.
ID P71497
AC P71497;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ClpB (Fragment).
GN Name=clpB;
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25416;
RX MEDLINE=97148974; PubMed=8995799;
RA Falah M., Gupta R.S.;
RT "Phylogenetic analysis of mycoplasmas based on Hsp70 sequences: cloning of the dnaK (hsp70) gene region of Mycoplasma capricolum.";
RL Int. J. Syst. Bacteriol. 47:38-45(1997).
DR EMBL; U51235; AAB09427.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001270; Chaperin clpA/B.
DR PRINTS; PR00300; CLPPTREASEA.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding.
FT NON TPR
SQ SEQUENCE 495 AA; 56358 MW; A194DD51FACDF8D0 CRC64;

Query Match      8.1%; Score 122; DB 2; Length 495;
Best Local Similarity 24.6%; Pred. No. 6.1;
Matches 82; Conservative 37; Mismatches 113; Indels 102; Gaps 16;

Qy 2 TSIAEQTVVVKSAIETADGALDLYNKILQDVIPKTFDETIKLSR 49
Db 178 TGIIVDKLISGEKRLNLDLLKKYVGQDQAIKAVTSAIMRSRSGKPKDPKPGISFLF 237
Qy 50 F-----KQEYSQBS-ASVLVGDIKVLLMDSQDKYFEATQTV-----YEWCGVWT 91
Db 238 FGPTGVGKTEVARSLADILFNSPKMIRLDMSEYNEKHSVAKLIGAPPYGVYEGGRILT 297

Query Match      8.0%; Score 121; DB 2; Length 2723;
Best Local Similarity 19.4%; Pred. No. 54;
Matches 60; Conservative 54; Mismatches 102; Indels 94; Gaps 9;

Qy 11 EVKSAIETADGALDLYNKILQDVIPKTFDETIKLSRPFQYSQBSASVLGVGDKVLLM 53
Db 1464 EELKGYIDSKSGYKDEVNKVT-----KTIENNKCELEKYKNEAINLNLKSEFPLKNKI 1517
Qy 54 --YSQASVLVGDIKVLLMDSQDKYFEATQTVYEWGCVVTTLLSAYILLPDEYNEKKA 111
Db 1518 IQTKDLEIITINIKKHKES-----ILOADKSEQKINT 1551
Qy 112 QKDILIRILDGKVKLNEAOKSLTSSQSFNNASGKLLALDSOLTNDSEKSSYFQSQVD 171
Db 1552 IKNEIQIKDDTTSN-DKSNKAIIGIQASLDKDKETKFLKI-----NDIRTKSNFLKETE 1605
Qy 172 RIRKEAYAGAAAGVAGPFGLIISYSIAAGVIEGKLIPELNLKTVQNTFTLSATVKQ 231
```

```
Db 1606 NIEQK-----ISNLSINSQEIK---LKNEDILNTLQKFFSLKDKQKN 1646
Qy 232 -----ANKDIDAAKLKLATEIAAIGBIKTETETFRFYVDYDMLMLSLK 276
Db 1647 IEDQKTELDNFDSEIENIESDVQKNKYEMGLIEKIKENADTNKQIESTK--ELIKP 1703
Qy 277 AAKKWINTCN 286
Db 1704 TIENLISSFN 1713
RESULT 16
Q74DE2 PRELIMINARY; PRT; 541 AA.
AC Q74DE2;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN Name=hylB; ORFNames=GSU1374;
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCA / ATCC 51573;
RX PubMed=14671304; DOI=10.1126/science.1088727;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
Davidson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
Weidman J.F., Khouri H.M., Feldblyum T.V., Utterback T.R.,
Van Aken S.E., Lovley D.R., Fraser C.M.;
RT "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments.";
RL Science 302:1967-1969(2003).
DR EMBL; AF017180; AAR34750.1; -.
DR TIGR; GSU1374; -.
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;
Query Match 7.9%; Score 119; DB 2; Length 541;
Best Local Similarity 23.5%; Pred. No. 11;
Matches 69; Conservative 44; Mismatches 103; Indels 78; Gaps 12;
Qy 19 TADGALDLYNKYLDQVTPWKTFDETIKELSRFKQYSQEAASVLVGGIKVLLMDSQDKYFE 78
Db 36 TANNGLD--TVYRDRVLP-----LKLKLIADMY-----AVNIVDVSHKVRNGNITWE 82
Qy 79 ATQTVYEWCGVVTQLLSAYILLFDEYNEKKSAQKDLIRILDGKVKLNEAQSLLTSS 138
Db 83 GRKSVEEAKKTIAEKLOAYLATNLAEEKKHLEAKPLIKVADATLER-----LASI 134
Qy 139 QSFNNAAGKLLALDSQLTNDSEKSYFQSQVD---RIRKEAY-----A 179
Db 135 LSAEDAEALTAFTVSELYPAIDPVSAKFSLSVDDQLKIAKQEYDHSGLYRASRTISLVA 194
Qy 180 GAAAGIVAGPFGLIISYIAAGVIEGKLIPELNRLK-----TVQ----- 219
Db 195 IIVGVLIAGTAGLITTSITGTPLAEG---VEVANRLAAGDLTVEVRAGGDETGQLMAM 251
Qy 220 -NFTSL-----SATVQKANKIDIAAKLKLAT---EIAA--IGEIKTETE 258
Db 252 GNMVTSRLHLIAEAISSHIGIASNQLHATSEQIATGSEVASQVGAVATASE 305
RESULT 17
AAR34750 PRELIMINARY; PRT; 541 AA.
ID AAR34750;
AC AAR34750;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN HYL B OR GSU1374.
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCA / ATCC 51573;
RX PubMed=14671304;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
Davidson T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
Weidman J.F., Khouri H.M., Feldblyum T.V., Utterback T.R.,
Van Aken S.E., Lovley D.R., Fraser C.M.;
RT "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments.";
RL Science 302:1967-1969(2003).
DR EMBL; AF017211; AAR34750.1; -.
DR TIGR; GSU1374; -.
SQ SEQUENCE 541 AA; 57093 MW; 3C9BEDAEAF7C3812 CRC64;
Query Match 7.9%; Score 119; DB 2; Length 541;
Best Local Similarity 23.5%; Pred. No. 11;
Matches 69; Conservative 44; Mismatches 103; Indels 78; Gaps 12;
Qy 19 TADGALDLYNKYLDQVTPWKTFDETIKELSRFKQYSQEAASVLVGGIKVLLMDSQDKYFE 78
Db 36 TANNGLD--TVYRDRVLP-----LKLKLIADMY-----AVNIVDVSHKVRNGNITWE 82
Qy 79 ATQTVYEWCGVVTQLLSAYILLFDEYNEKKSAQKDLIRILDGKVKLNEAQSLLTSS 138
Db 83 GRKSVEEAKKTIAEKLOAYLATNLAEEKKHLEAKPLIKVADATLER-----LASI 134
Qy 139 QSFNNAAGKLLALDSQLTNDSEKSYFQSQVD---RIRKEAY-----A 179
Db 135 LSAEDAEALTAFTVSELYPAIDPVSAKFSLSVDDQLKIAKQEYDHSGLYRASRTISLVA 194
Qy 180 GAAAGIVAGPFGLIISYIAAGVIEGKLIPELNRLK-----TVQ----- 219
Db 195 IIVGVLIAGTAGLITTSITGTPLAEG---VEVANRLAAGDLTVEVRAGGDETGQLMAM 251
Qy 220 -NFTSL-----SATVQKANKIDIAAKLKLAT---EIAA--IGEIKTETE 258
Db 252 GNMVTSRLHLIAEAISSHIGIASNQLHATSEQIATGSEVASQVGAVATASE 305
RESULT 18
Q9SZK7 PRELIMINARY; PRT; 1496 AA.
ID Q9SZK7;
AC Q9SZK7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Hypothetical protein F20D10.190 (Hypothetical protein At4g38070).
GN Name=F20D10.190; Synonyms=AT4g38070;
```

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Schueller C.;
 RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035538; CAB37547.1; -
 DR EMBL; AL161592; CAB80472.1; -
 DR PIR; T05634; T05634.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PF00010; HLH; 1.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS00888; HLH; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1496 AA; 173346 MW; 03BBB75DCD2B03FE CRC64;
 Query Match 7.8%; Score 118; DB 2; Length 1496;
 Best Local Similarity 19.6%; Pred. No. 42;
 Matches 64; Conservative 67; Mismatches 122; Indels 74; Gaps 13;
 QY 4 IFRAQTEVVKSAIETADGALDLYNK--YLDQVTPWKT-----PDETIKELSRP----KQ 52
 DB 153 IEAEKTVKMKRGRDVRVVKVEEKSQVBEKLNKKEQFKHLEAYEKLKFLKDSKK 212
 QY 53 EYQEAASVLVGDIVL--LMSODKYFEATQTVVEMC-GVVTQLLSAYILLFDEYNEKKA 109
 DB 213 EWEBEKSLDDEIYSLQTKDVSRISEDLQKLCMCNGALQ-----BETRKHLL 263
 QY 110 SAQKDILIRLDGVKKLNEAQSLLTSFSSFNASGKLALDLSQLTDFSEKSSYFQ-- 167
 DB 264 EIQVSEPKAKYEDAFASQDARTQL-----DDLACKRDWEVAELRQTLQTLQKMDAYPEM 316
 QY 168 ----SQVDRIRKEAYAGAAAGIAGVAPGLIISYIAAGVIEG---KLIPELNRLKTVQN 220
 DB 317 KYENGKLEQENRELLGSLK-----ELQBATIQGSGNSALSCLKNKNFRLEN 362
 QY 221 FFTSLSAT-----VKOANKDIDAAKLATIAAIGETKTE-----TTRFYV 264
 DB 363 IHRKCSANLAKESKEWSEKQVKEEINDYKLQSKAEALKKEVELENCRSSTAKMRL 422
 QY 265 DYDDLMLSLKGAQKMNCTNEYQQR 291
 DB 423 QYBEISIMEL-----VLSRTVSEASQR 444
 RESULT 19
 ID Q6BFD6 PRELIMINARY; PRT; 1877 AA.
 AC Q6BFD6;
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=PTMB.434;
 OS Paramesidium tetraurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;

OC Paramesidium.
 OX NCBI_TaxID=5888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Stock d4-2;
 RA Zagulski M., Nowak J.K., Le Mouel A., Nowacki M., Migdalski A.,
 RA Gromadka R., Noel B., Blanc I., Dessen P., Wincker P., Kellier A.M.,
 RA Cohen J., Meyer E., Sperling L.;
 RA "high coding density on the largest Paramesidium tetraurelia somatic
 RT chromosome.";
 RL Curr. Biol. 0:0-0(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Stock d4-2;
 RA Nowak J.K., Migdalski A., Gromadka R., Zagulski M.;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CR548612; CAH03635.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1877 AA; 220626 MW; F68260FE840D2796 CRC64;
 Query Match 7.8%; Score 118; DB 2; Length 1877;
 Best Local Similarity 20.3%; Pred. No. 55;
 Matches 66; Conservative 64; Mismatches 135; Indels 60; Gaps 13;
 QY 7 EQTVEVVKSAIETADGALDLYNKYLDQVTPWKTDETIKELSRFKOEYSQEAASVYVGDIK 66
 DB 398 EQNKDILQHVANS-----FNKFLQEAENEQNAVFPKLLYNFFMEHDMHMLLIGQVS 450
 QY 67 VLIMDSQDKYFEATQTVY----EWCQVVTQLLSAYILLFDEYNEKKAQKQILIRL-- 120
 DB 451 FNLI-----KVETNHKKYPIVKSCKRLLESISFLLIVVNEVSKVYNTYINFLFSILIT 506
 QY 121 -DGVKK-----LNEAQSLLTSQSFN-----NASGKL-----ALDSQL 155
 DB 507 KEDNFEKSTIEFTLLRGAANLMTCLSKFSPEELKQSPGSTLNTLLVMEQLPQDLKQL 566
 QY 156 TNDPSEKSSYFQSQVDRIK--EAYAGAAAGIAGVAPGLIISY-SIAAGVIEGKLIPELN 212
 DB 567 VGP-AEMSNY----DAIEFGQLYQNLSDVLDGDEQIRKVFESIKSLIKGURFKQ-N 620
 QY 213 NRLKTVQNFTSLSATVVKQANKDIDAAKLATIAAIGETKTEETETTRFYVDYDDLMLSL 272
 DB 621 KEVQLPKWFTSTIKIENPSISLLAQSMIEILISEKVD-----IIVENLKL 671
 QY 273 LLKGAQKMNCTNEYQQRHGKTL 297
 DB 672 IIBESRTKMLKNQIVGQDYTKLTL 696
 RESULT 20
 ID Q7N5U5 PRELIMINARY; PRT; 577 AA.
 AC Q7N5U5;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Similar to putative phage protein.
 GN OrderedLocusNames=plui1837;
 OS Photobacterium luminescens (subsp. laumondii).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Photobacterium.
 OX NCBI_TaxID=141679;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TT01;
 RX MEDLINE=22957627; PubMed=14528314;
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
 RA Taouit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
 RA Dassa E., Deroose R., Deruelle S., Freysinet G., Gaudrault S.,
 RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
 RT "The genome sequence of the entomopathogenic bacterium Photobacterium luminescens.";

```

RL  Nat. Biotechnol. 21:1307-1313 (2003).
DR  EMBL; BX571865; CAE14130.1; -.
DR  Photolist; plu1837; -.
KW  Complete proteome.
SQ  SEQUENCE 577 AA; 65287 MW; B7F86EE4B34BEDE1 CRC64;

Query Match          7.8%; Score 117.5; DB 2; Length 577;
Best Local Similarity 19.2%; Pred. No. 15;
Matches 73; Conservative 56; Mismatches 142; Indels 109; Gaps 12;

QY  17 IETADGALDLYNKYLDQV---IPWKTFF-----DETI 44
DB  129 LETPPAYVFLPFIIDQIKSWSSPNSFENLGOYSNWKGPLIKYFTGYLKPEHPDIEENI 188
QY  45 KLSRFRKQEQSASVILVGDIKVL-----LMSODK 75
DB  189 YEYSEIKESAHKTEKQSAVEVIVDNTVDSPTLDNNDFIKQNEIKNELYLDIDYQTR 248
QY  76 YFEATQTVYEWGVVQTLLSAYILLPDEYNEKKA-----SAQKOIL----- 116
DB  249 LYDAQATI---TSNIYDLESQYALAITSAHELEADYKFAVESIPTDYLECPGLCTLHDNS 305
QY  117 ----IRILDGVKKLINEAKQLLTSSQSFNNASCKLLALDSQLTNDFSEKSYFQSOVDRI 172
DB  306 LPNRALLSEKDSLNEA-----NSIASKIVELKSSL-NSLNEDAQFITNEIER 353
QY  173 IRKEAYAGAAA---GIVAGPGLIISYIAGV-----IEGKLIPELNNRLKTVQNFPTS 224
DB  354 INNKYITDDDAVKKGLIAQVITLSTENVSKNIQIKINDENLNISKANSISKELKDDQKK 413
QY  225 LSATVKQANKOIDAAKLAKEIAAIGETETETTRFYVDYDMLSLKGAAKMI-N 283
DB  414 LUST-KEKEELNSPMSKLLANIELALGSTGINSKVSPTDYKQLGGGAEEAARGLLAY 472
QY  284 TCNEYQORHGKTKLTFEVPDV 303
DB  473 QLSVLQIHTGAKTCVVPFV 492

RESULT 21
Q7Q004 PRELIMINARY; PRT; 1224 AA.
AC Q7Q004;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE GLP_227_22033_18359.
OS Giardia lamblia ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=184922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB C6; McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RA Morrison H.G., Sogin M.L.;
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACB01000157; EAA37104.1; -.
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; Ank; 5.
CC PROSITE; PS50088; ANK_REPEAT; 3.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC ANK repeat.
SQ SEQUENCE 1224 AA; 136345 MW; 8BBE76BEC6B17E9C CRC64;

Query Match          7.8%; Score 117.5; DB 2; Length 1224;
Best Local Similarity 19.2%; Pred. No. 36;
Matches 65; Conservative 57; Mismatches 126; Indels 91; Gaps 10;

QY  2 TSIFAEQTVVEVVKSAIETADGALDLYNKYL-----DQV-----IPW 37

```

```

DB  876 TIIDLEQSLDNAQVSTSDADISRLNDELSTLKOQSSKCLEEANSKIDLLKAELEE 935
QY  38 KTFD-----ETIKELSRFKQEQSASVILVGDIKVLLMDSQDKYFEATQTVYEWCGV 89
DB  936 KTIESASNPDLSSIPSLARSTSESTQNLGELSDELRELLDSERRRIKKLN-----NK 989
QY  90 VTQLLSAYILLPDEYNEKKASAQK-----DILIRILDGVKKLNE 129
DB  990 ITALMTEKNLNDQINAMKTSNAKNLPLPAAPVLEEPSLATGTDDVRISOLEQALKDANE 1049
QY  130 ----AQKSLLTSSQSFNNASCKLLALDSQLTNDFSEKSYFQSOVDRIKEAVAGAAAGI 185
DB  1050 RADAKERELTDALORIENLGAQISPLEAEUSS-IQDRLSATGEVIEKLRANQQ----- 1102
QY  186 VAGPPGLIISYIAAGVIEGKLIPELNNRLKTVQNFPTSISATVKQANKOID----- 237
DB  1103 -----LNQIAGAAGQSEBELITSLTRELEGLTENSQNLNAVKEDLKQIEDKVKELAI 1154
QY  238 ----AAKILKATEI--AAIGETETETTRFYVDYDML 270
DB  1155 ERQQAETLKLQRIQIPDKIASYSDHVEKPRSVNESDDVM 1193

RESULT 22
HLV1_ECOLI STANDARD; PRT; 1023 AA.
ID HLY1_ECOLI
AC P09983;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin, chromosomal.
GN Name=hlyA;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J96 / Serotype O4;
RC MEDLINE=85234404; PubMed=3991743;
RX Felmler T., Pellett S., Welch R.A.;
RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";
RL J. Bacteriol. 163:94-105 (1985).
RN [2]
RP SEQUENCE OF 1-44 FROM N.A.
RC STRAIN=2001;
RC MEDLINE=85258115; PubMed=3894051;
RX Nicaud J.-M., Mackman N., Gray L., Holland I.B.;
RA "Characterisation of HlyC and mechanism of activation and secretion of
RT haemolysin from E. coli 2001.";
RL FEBS Lett. 187:339-344 (1985).
CC -1- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: The Gly-rich region is probably involved in binding
CC calcium, which is required for target cell-binding or cytolytic
CC activity.
CC -1- DOMAIN: The three transmembrane domains are believed to be
CC involved in pore formation by the cytotoxin.
CC -1- PTM: Palmitoylated by hlyC. The toxin only becomes active when
CC modified.
CC -1- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly
CC by strains causing extraintestinal infections, such as those of
CC the urinary tract.
CC -1- SIMILARITY: Belongs to the RTX prokaryotic toxin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
CC EMBL; M10133; AAA23975.1; --
DR EMBL; X02768; CAA26546.1; --
DR PIR; A24433; LRECA.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR003995; RtxA.
DR InterPro; IPR011049; Serralyen_like_C.
DR Pfam; PF00353; HemolysinCabin; 6.
DR Pfam; PF02382; RTX; 1.
DR PRINTS; PR00313; CABNDNGRPT.
DR PRINTS; PR01488; RTXTOXINA.
DR PROSITE; PS00330; HEMOLYSIN CALCIUM; 4.
KW Calcium; Cytolysis; Hemolysis; Lipoprotein; Palmitate; Repeat; Toxin;
KW Transmembrane.
FT TRANSMEM 237 259 Potential.
FT TRANSMEM 267 326 Potential.
FT TRANSMEM 364 410 Potential.
FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.
FT REPEAT 723 728 1.
FT REPEAT 732 737 2.
FT REPEAT 741 746 3.
FT REPEAT 750 755 4.
FT REPEAT 759 764 5.
FT REPEAT 768 773 6.
FT REPEAT 777 782 7.
FT REPEAT 786 791 8.
FT REPEAT 795 799 9.
FT REPEAT 806 812 10.
FT REPEAT 816 821 11.
FT REPEAT 825 830 12.
FT REPEAT 834 839 13.
FT REPEAT 843 848 14.
FT REPEAT 855 860 15.
FT REPEAT 864 869 16.
FT LIPID 563 563 N(6)-palmitoyl lysine (By similarity).
FT LIPID 689 689 N(6)-palmitoyl lysine (By similarity).
FT VARIANT 6 6 A -> T (in strain 2001).
SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;

Query Match 7.7%; Score 116; DB 1; Length 1023;
Best Local Similarity 20.5%; Pred. No. 36;
Matches 69; Conservative 59; Mismatches 111; Indels 98; Gaps 11;
QY 2 TSFAEQTVWVK-----SAIETADGALDLYNKYLQVPIPKTPTDETIKELSRFKQE 53
DB 155 TALSSMKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207
QY 54 YSQEASVL-----VGDYKVLMDSDQKYFEATQTVYEWGVVTVQLLSAYILL-PDEY 104
DB 208 LNKLGSLVNTKHLNGVGN-KLQNLPLNDNIGAGLDTV---SGTSAISAFILSNADAD 263
QY 105 NEKKASAKQILIRLDD---GVKLNKAEQKSLTSSQSFNNASGKLALDLSQITNDFSE 161
DB 264 TGTAAAGVELTKVLGNVGKISQYIIAQAAGLST----- 302
QY 162 KSSYFOSQVDRIRKEAYAGAAAGIVAGPFGGLIIS---YSIAAGVIEGKLIPELNNRKT 217
DB 303 -----AAAGLIASVTVTLAISPLSFLSIADKFKRANKIEYSQFFKK 344
QY 218 VQNPFTSLSTVQKANKIDAAKLAETAAIGETETTRFYVDYDDLMLSLKGA 277
DB 345 LGYDGDLSLAFFHKTGAIDASLTRISTVLASVSSGISAAATTSLV---GAPVSAIVGA 400
QY 278 -----AKMINTCNEYQQRHGK 294
DB 401 VTGIISGILEASKOAMPEHVASKMADVIAEWKKGK 437

RESULT 23
Q7XEH4
ID Q7XEH4 PRELIMINARY; PRT; 2033 AA.

AC Q7XEH4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative centromere protein.
GN ORFNames=OSJNBa0075H14.2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA The Rice Chromosome 10 Sequencing Consortium;
RT "in-depth view of structure, activity, and evolution of rice
RL chromosome 10.";
RL Science 300:1566-1569 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017095; AAP53815.1; --
DR Gramene; Q7XEH4; --
DR InterPro; IPR001917; Aminotrans II.
DR PROSITE; PS00599; AA_TRANSFER CLASS 2; UNKNOWN 1.
SQ SEQUENCE 2033 AA; 231439 MW; ACFD716945B5C462 CRC64;
Query Match 7.7%; Score 116; DB 2; Length 2033;
Best Local Similarity 18.7%; Pred. No. 82;
Matches 67; Conservative 62; Mismatches 106; Indels 124; Gaps 11;
QY 7 EQTVVVKSAIE-----TADGA-----LDLYNKYLQVPIPKTPTDETIKELSRFKQEYSQEA 58
DB 477 EQKVQMLQDLEQKQEQADSAHAQLQDECNH-----TQTEADLHFRKLNLSQ-- 524
QY 59 SVLVGDIKVLMDSDQKYFEATQTVYEWGVTVQLLSAYILLPDEYNEKKA-----SAQK 113
DB 525 -----LEEVIKLTENLDRSTKSELENAKLDLENTRELKSTLDLNLSEK 571
QY 114 DILIRILDGKVLKNEAQSLSLTSSQSFNNASGKLALDLSQLT-----NDFS 160
DB 572 DAVLQQQSLAKISELEQLSKTQLELNKSEKQMLLELEITQKSEMSDSLTLSLKDET 631
QY 161 EK-----SSYFOSQ---VDRIRKEAYAGAAAGIVAGPFGGLIISYIAAGVIEGK 206
DB 632 EKVQVQAEATSLMSMESMYSQSQEEVNRHLHLE----- 661
QY 207 LIPELNNRKTQNPFTSLSATVQKANKIDAAKLAETAAIGETETTRFYV-- 264
DB 662 -IEKLNFKLNELENLSSELNLTILLNAEKDADTLKRNQKSLVRLSLESLSKLAQLEK 720
QY 265 -----DYDDLMLSLKGAKK-----MINTCNEYQQRHGKKT 297
DB 721 IEKGVQMLEQELKHKEEVDLSLQISIQDEAHKSEGEAALLAMTNLNSQEEVNRLLT 779
RESULT 24
SR68 CAEEL
ID SR68 CAEEL STANDARD; PRT; 622 AA.
AC Q2082;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable signal recognition particle 68 kDa protein (SRP68).
GN ORFNames=F55C5.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Harris B.;

```
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: signal-recognition particle assembly has a crucial role
CC in targeting secretory proteins to the rough endoplasmic reticulum
CC membrane. SRP68 binds the 7S RNA, SRP72 binds to this complex
CC subsequently. This ribonucleoprotein complex might interact
CC directly with the docking protein in the ER membrane and possibly
CC participate in the elongation arrest function (by similarity).
CC -1- SUBUNIT: Signal recognition particle consists of a 7S RNA molecule
CC of 300 nucleotides and six protein subunits: SRP72, SRP68, SRP54,
CC SRP19, SRP14 and SRP9 (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the SRP68 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z78198; CAB01573.1; -.
CC FIP: T22716; T22716
CC WormPep: F55G5.8; C820875.
CC InterPro: IPR008941; TPR-like.
CC KW Hypothetical protein; Ribonucleoprotein; RNA-binding;
CC Signal recognition particle.
CC SQ SEQUENCE 622 AA; 70574 MW; A7B8808E46169636 CRC64;
CC
CC Query Match 7.6%; Score 115.5; DB 1; Length 622;
CC Best Local Similarity 21.2%; Pred. No. 22;
CC Matches 68; Conservative 56; Mismatches 109; Indels 87; Gaps 15;
CC
CC QY 41 DETIKELSRFKQEQYSQASVLVGDIVKLLMSQDKYFEATQTVVWCGVVTQLLSAY--- 97
CC Db 251 DKLISEM-RASATSAEVVTVIENGGAQSTVDDE-----KAKQVQEWKQTEVELAQCOPTK 304
CC
CC QY 98 --ILLFDEYNEKASAKQKIDILRLDDGVKKLNE-AQKSLTSSQSF-----NNASGKL 148
CC Db 305 EKMALEP-----EKATADTRDAIDRIISIRKSSSENADTVVLSQIKAYLEFLKMGNTASRY 360
CC
CC QY 149 LALDSQLTNDFSEKSYFOS-----QVDRIRKEAYAG 180
CC Db 361 LAI---IDNTKSEKSKPKQDLRLLYDSVIEIYKVAEIPGADHDKNLIQAPEVKVEYRA 417
CC
CC QY 181 AAGIVAGPFLGIISYIAAGVI-----EGKLIPELNRLKTVQNFFTLSATVK 230
CC Db 418 FRCFYMASSYSALHKYSEAAALFDRTVSRVQDAEGL-----KKLKSSFITNETQSSIN 472
CC
CC QY 231 QANKDIDAAK--LKLATEIAAIGEIKTETETTRFFVYDDMLSLKGAAKKMTNCFY 288
CC Db 473 ELRSEVESAKVTVAARLASAGDVKTDSLAKI-ID-----KRLLETVNEW 519
CC
CC QY 289 QQ---RHGKKTLEFPDVAS 305
CC Db 520 QWDVRSNLKDKKTIIP-VAS 538
CC
CC RESULT 25
CC Q97WH8 PRELIMINARY; PRT; 587 AA.
CC AC Q97WH8;
CC DT 01-OCT-2001 (TrEMBLrel. 18, Created)
CC DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
CC DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
CC DE BPS2 protein homolog (Bps2).
CC GN Name=Bps2; OrderedLocusNames=SSO2241;
CC OS Sulfolobus solfataricus.
CC OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
CC OC Sulfolobus.
CC OX NCBI_TaxID=2287;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC
CC STRAIN=ATCC 35092 / DSM 1617 / P2;
CC MEDLINE=21332296; PubMed=11427726;
CC RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
CC RA Aweez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
CC RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,
CC RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
CC RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
CC RA Charlbois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
CC RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
CC RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
CC RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC DR EMBL: AE006828; AAK42408.1; -.
CC DR PIR: A90394; A90394.
CC KW Complete proteome.
CC SQ SEQUENCE 587 AA; 68436 MW; B90DBC1E19C05E86 CRC64;
CC
CC Query Match 7.6%; Score 115; DB 2; Length 587;
CC Best Local Similarity 20.3%; Pred. No. 22;
CC Matches 61; Conservative 66; Mismatches 103; Indels 70; Gaps 14;
CC
CC QY 10 VEVVKSALETADGALDLYNKYLQVVPKTFDETI--KELSRFKQEQYSQASVLVGDIVK 67
CC Db 42 ISLTSSIK-----AEDLLNVFADSGYVEAELDNKLYRRIRKIRNGELGEKNLMDDDRA 97
CC
CC QY 68 LLMSQDKYFEATQTVVWCGVVTQLLSA-----YILLFDEYNEKASAKQKIDILRLD 121
CC Db 98 LLL-----TYFSPENR-----LVQTILSGDGNVWFSTTSKINEIK--AKKEELQKLT 145
CC
CC QY 122 DGVKKLNEAQKSLTSSQSFNN--ASGKLALDSQLTNDFSEKSSYFOSQVDRIRKEAY 178
CC Db 146 AEINARDELQK-----KYNIREIQAKIRAI-----EIDKLEKE-- 181
CC
CC QY 179 AGAAAGIVAGPFLGIISYIAAGVIEGKLIPELNRLKTVQNFFTLSATVKQAKDIDA 238
CC Db 182 RESSSNIVAK-----TYYTIT--LFRONKINEILNKIKVKKDELANLEFALKIEEELQN 234
CC
CC QY 239 AKLKATEIAAIGEIKTETETTRFFVYDD-----LMLSLKGAAKK-----INTCN 286
CC Db 235 KESKVSPIKTLQEKEMEEINEKCLKTNDRSELSELKLVRLVEEVNEDRHHLDTCN 294
CC
CC RESULT 26
CC O64067 PRELIMINARY; PRT; 478 AA.
CC AC O64067;
CC DT 01-AUG-1998 (TrEMBLrel. 07, Created)
CC DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
CC DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
CC DE Hypothetical protein yonb.
CC GN Name=yonb;
CC OS Bacteriophage SPBc2.
CC OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
CC OX NCBI_TaxID=66797;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mavel C.,
CC RA Karamata D.;
CC RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC DR EMBL: AF020713; AAC13027.1; -.
CC DR PIR: T12818; T12818.
CC KW Hypothetical protein.
CC SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;
CC
CC Query Match 7.6%; Score 114.5; DB 2; Length 478;
CC Best Local Similarity 22.6%; Pred. No. 19;
CC Matches 57; Conservative 41; Mismatches 73; Indels 81; Gaps 12;
CC
CC QY 25 DLYNKY-IDQVVPKTFDETIKELSRFKQEQYSQASVLVGDIVKLLMSQDKYFEATQV 83
CC Db 268 DVIDTYFTVNVYWS--DENSVD-KYPKFNTR-----TGDVTSIDFDSKTEVF---MT 315
CC
CC QY 84 YEWCGVVTQLLSAYILLFDEYNEKASAKQKIDILRLDDGVKKLNEAQKSLTSSQSFNN 143
CC RP SEQUENCE FROM N.A.
```



```

GN OrderedLocusNames=GBAA_px01_0124;
OS Bacillus anthracis.
OG Plasmid virulence plasmid px01, and Plasmid px01.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RX MEDLINE=9445483; PubMed=10515943;
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
RA Keim P., Koehler T.M., Lanke G., Kumano S., Mahillon J., Manter D.,
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.;
RT "Sequence and organization of px01, the large Bacillus anthracis
RT plasmid harboring the anthrax toxin genes.";
RL J. Bacteriol. 181:6509-6515(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581; PLASMID=px01;
RA Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065404; AAD32394.1; -.
DR EMBL; AE017336; AAT28865.2; -.
DR PIR; B59102; B59102.
DR TIGR; GBAA_px01_0124; -.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH DOMAIN; UNKNOWN_1.
KW Complete proteome; Plasmid.
SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;

Query Match 7.6%; Score 114.5; DB 2; Length 652;
Best Local Similarity 18.6%; Pred. No. 27;
Matches 58; Conservative 66; Mismatches 106; Indels 81; Gaps 10;

Qy 25 DLYNKYLQVVPKPTFDTIKELSRFKQVYQSASVL-----VGDIKVLMDSQDKYFE 78
Db 237 DVINQKIDE---FDKLSQRKDLERMLELNQKLSQLKQSPQLQDLNKLKESQSRLLE 293
Qy 79 ATQ-----TVYEWCGVVTQLLSAYILLFPEYNE--KKASAQKDILIRILDDGVKK 126
Db 294 LNKDSNRLEINSEIKKLNDRKAEILLSIMELIKQSQSEPKKIKNEKDDLKKKREDLINR 353
Qy 127 LNEAQK-----SLTSSQSFNNAAGKLLALDSQLTNDPSEKSSYFQ---SQV 170
Db 354 IAESKELAKKAEINTKLVELFKVQBALNKKSGQYLYYINKLDNELRELADKYKNSDNKI 413
Qy 171 DRIRKEAYAGAAGIVAGPGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTLSATVK 230
Db 414 SRLKNH-----IGBYNKQLEKIEN-----ELE 435
Qy 231 QANKDIDAACKLAT-EIAAIGELTKETETTRFVVDYDDL-----MLSLKGAACKMINT 284
Db 436 ECNKKIDNTKKQLAEFDKSNKKQOELESELVQLNKKIDELGKHKHQRHELEASQKKALDE 495
Qy 285 CNEYQQRHGKK 295
Db 496 AKEINKKLAEK 506

RESULT 32
ID Q7RM79 PRELIMINARY; PRT; 779 AA.
AC Q7RM79;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE GTPase of unknown function, putative.
GN Names=PY02306;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalton S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,
RA Florens L., Yates P.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., O.R.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White M.J.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an

```

Query Match	7.6%; Score 114.5; DB 2; Length 1885;
Best Local Similarity	20.7%; Pred. No. 33; Indels 132; Gaps 14;
Matches	76; Conservative 51; Mismatches 109; Indels 132; Gaps 14;
QY	40 PDETIKELSRFKQESYQBSASVL-----VGDIKVLMDSQDKYFEA 79
Db	942 FDSQODNQSIQSNQLESTLTLAQSENORLLTENKQFITSLNEIKSLFNSIQOQK-ET 1000
QY	80 TQTVYWGCVVQ-----LLSAYILLFDEYNEKSAQKDILIRLLDGVKKLNEAQSLLT 136
Db	1001 IQLEFEFOKENQFDSLJTNYNQLFKYNDLATSNESNRL--EFDQFKDSNQSIQSLES 1058
QY	137 SSSQFNASGKLL-----ALDSOLT-----DFSEKSSYFQSQVD 171
Db	1059 LERSLKSNDNLQSSLLKQLESIEKQKODQILPIQLELESKKCELSLSSQFSQTK 1118
QY	172 RIRKEAYAGAAAGIVAGPGLIIS---YSIAAGVIEGKI---PELNN----- 213
Db	1119 QVTVQ-----LLISVDQYKISTNKLESQSRNREINNLKKAIEINALK 1162
QY	214 -----RLKTVQNFSTSLSAVTVKQANKD---IDAAK 240
Db	1163 EENISLKQDITLKKAPKSQTDREXDMIKLEKLEKREFAIDAKLQAIQDKQTIOSEK 1222
QY	241 LKLAETIAAIGEIKTETETTRFYVDYDMLSLKLG-----AAKWMINTCNEYQORH 292
Db	1223 QSLREIKDKRSHTSTET-----ELDKLKKTHLAADVKSDFIALNKSVELTKSQQL 1277
QY	293 GKXTLFEV 300
Db	1278 -KSTIIEL 1284

RESULT 34

Q9H6Q7 PRELIMINARY; PRT; 720 AA.

AC Q9H6Q7

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical protein FLJ21979.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_taxid=9606;

RN [1]

RA SEQUENCE FROM N.A.

RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK025632; BAB15196.1; --

FT NON_TER 720

SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953DOB CRC64;

Query Match 7.5%; Score 114; DB 2; Length 720;

Best Local Similarity 20.7%; Pred. No. 33;

Matches 62; Conservative 56; Mismatches 111; Indels 70; Gaps 10;

QY	26 LYNKVLQDVIWKTFDETIELSRFQSYQBSASVLVGD----IKVLMDSQDKYFEATQ 81
Db	457 LYKEYLESEKETWKTESKIIEKKRKLQDOVQDAIKVKEYNNLLNALQMSD----EMKK 512
QY	82 TVYWGCVVQTLLSAYILLFDBYNPKKSAQKDILIRLLDGVKKLNEAQSLLTSSQS 141
Db	513 ILAENSRIKTVL-----QVNEKSLIRQVTLTVE--LERQLRKENEKQN----- 554
QY	142 NNASGKLLALDSQLTNDPSEKSSYFQSQVDRIKREYAGAAAGIVAGPGLIISYIAG 201
Db	555 -----ELLGWEAEV-----CEKIGCQLQ-----RPEKWAIFKIAA-----LQKVDNDSVLS 595

KW ATP-binding; Microtubule; Motor protein.
SQ SEQUENCE 1885 AA; 218215 MW; 15686ABD3B007EC7 CRC64;

Query Match 7.6%; Score 114.5; DB 2; Length 1885;
Best Local Similarity 20.7%; Pred. No. 33;
Matches 76; Conservative 51; Mismatches 109; Indels 132; Gaps 14;

QY 40 FDETIKLSRFKQEYSQAASVL-----VGDIKVLLMDSQKYFEA 79
DB 942 FDQQDNQSIQSINOLESTTLTLAGSENORLLTENKOFITSLNEIKSLFNSIQOQK-ET 1000

QY 80 TOTVYWCVVVTQ---LTSAYILLDFEYNKKASAQKDILIRILDGDKVKLNKAQSKILT 136
DB 1001 IQLEFFEQEKENQFDLLTNQLFSKYNDLATSNRL--EPDPFKDSNQISQLES 1058

QY 137 SSSEFNNAACKLI-----ALDSOLTN-----DFSEKSSVFQSOVD 171
DB 1059 LERSLKSENDNLLQQSLLKSQLESIEKQKQDLPIQLELESKKCELSKLSSFSEQTK 1118

QY 172 RIRKEVAGAAAGIVAGPGLIIS---YSIAGVLEGKI---PELNN-----213
DB 1119 QVTQ-----LLISVDQYKISTNKUESQISDRNEINNLLKATRNALK 1162

QY 214 -----RLKTQNFTFSJSA TVKQANKD----IDAK 240
DB 1163 BENISLDQLTLYKKAPKQTREKDWIKELKEKENAFDAIDLKQAIODKQTOISEK 1222

QY 241 LKLAETAAGAIGETETTRTFVDDMLSLKG-----RAKWMINTCNEYQQRH 292
DB 1223 QSLEREINKLRSHSTSET-----ELDCLKTHAADVKSFIALNKSVEILTKSQBL 1277

QY 293 GKXLFV 300 : : : : :
DB 1278 -KSTIIEL 1284 : : : : :

RESULT 34
Q9H6Q7 PRELIMINARY; PRT; 720 AA.

ID Q9H6Q7 AC Q9H6Q7; 2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK025632; BAB15196.1; --
FT NON TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match 7.5%; Score 114; DB 2; Length 720;
Best Local Similarity 20.7%; Pred. No. 33;
Matches 62; Conservative 56; Mismatches 111; Indels 70; Gaps 10;

QY 26 LYNYLDQVIPWKTPDETTELKSRPKOEYSQEA SVLVGD----IKVLMDSQDYFAEQ 81
DB 457 LYKEYESEKETWTESKTKEEKKLEDQVQDAIKVXEYNNLLNALQMSD-----EMKK 512

QY 82 TVEWCCVTTQLLSAVILLDFEYNKKASAQKDILIRILDGDKVKLNKAQSKILTSSOSF 141
DB 513 ILAENSRIKTVL-----QVNKSLRQTYTLIVE-LEROLRKENEKQKN-----554

QY 142 NNASKLLALDSLTDNFSEKSYFQSDVRIRKRAYAGAAAAGIVAGPFGLIISVIAG 201
DB 555 ----ELLSMEEAV-----CEKICLO-----RKFMAIFKIAA-----LQKVIDNSVLS 595

EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.

CC EMBL; AABL0100631; EAA21743.1; --
DR GO; GO:0005525; F:GTP binding; IEA.
DR InterPro; IPR005289; GTP-binding.
DR InterPro; IPR002917; MMR HSR1.
DR Pfam; PF01926; MMR HSR1; 1.
DR TIGRFAMs; TIGR00650; MG442; 1.
SQ SEQUENCE 779 AA; 92336 MW; 7F51893C396B0B52 CRC64;

Query Match 7.6%; Score 114.5; DB 2; Length 779;
Best Local Similarity 20.5%; Pred. No. 33;
Matches 71; Conservative 51; Mismatches 128; Indels 97; Gaps 14;

QY 24 LDLYNKVLDQVIPWKTPDETTELKSRPKOEYSQEA SVLVGDIKVLLMDSQDYFEATQTV 83
DB 92 LGYKGEVENTEFFQTFDEFNV--SRSEDYEQ---LEAGKLI-TENTNTYKDIRKI 145

QY 84 YEWCGVVTQLLSAYILLDFEYNKKASAQKD-----ILRILDD-----122
DB 146 ME-----KSEYKFLKKEBEEDSKISHDVHENLKVTFVKKNINDNFLTQY 193

QY 123 ---GVKKLNKAQSKLTSQSFNNAACKLIADLSQLTND-----FSEKSSYP-----QSQVD 171
DB 194 SYMNKLKEKSELKSVFNNGSTTSNDNIENNNDQVCLDEKSEKYSNDEKSKLR 253

QY 172 RIRKEVAGAAAGIVAGPGLIISVISYAAGVTEGKIIPNLN-----NELKTQVQFFT 223
DB 254 EINKPGDAENA-----LEFYKGSNKENKLIBEVNFPKPQKGYAEKRISTFSD 302

QY 224 SLSATVQKANK-----DIDAAXKLKATEIAAIGEIKETETTRTFVDDMLSLLL 274
DB 303 QMFANVKIENKLLIKEEYDWDGGKLEKEGEEVISNYYNDEGTFLFLKNVSVLELF 362

QY 275 KGAAAKMI-----NTCNE-----YQHGHGKTLFFEVPD 302
DB 363 KENVKASLIKVKKENIKNLLENLLCGRVKHWFHPKPMKRIIMKIPD 409

RESULT 33
Q869B8 PRELIMINARY; PRT; 1895 AA.

ID Q869B8 AC Q869B8; 2001 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyosellida; Dictyosteliaceae.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sungang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from Dictyostelium discoideum."
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Suyana E., Sutoh K.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB102780; BAC56912.1; --
DR HSP; F33173; 1161.
DR DictyBase; DDB0191404; kenD.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PRO0380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PSS0067; KINESIN MOTOR DOMAIN2; 1.

```
QY 202 VIE-----GKLIPELNRLKTVQNFSTLSATVQKANKDIDAALK 242
Db 596 EELANKQYNELTAKYRDILQKNMLVQRTSNLEHLECNISLKEQVESINKELEITKEK 655
QY 243 LAT-----BIAAIGEIKETETTRFRFVVDVDDMLSLKGAAGKMWINTCNEYOORHGK 295
Db 656 LHTEQAEQWETKLGNESSMDKAKSITNSDIVSISKKKKKKKKKKKKKKKKKKKK 714

RESULT 35
Q8CEIO PRELIMINARY; PRT; 1189 AA.
AC Q8CEIO
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Chromosome segregation SMC protein.
GN OrderedLocuNames=SE0909;
OS Staphylococcus epidermidis
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
  Qian Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
  Yan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
  Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
DR HESB; AE016747; AA004506.1; -.
DR HSPB; P58301; 1P2T.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0005525; F:GTP binding; IEA.
DR GO; GO:0007059; P:chromosome segregation; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005289; CTP-binding.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR010935; SMC_hinge.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF06470; SMC_hinge; 1.
DR Pfam; PF02463; SMC_N; 1.
DR TIGRFAMs; TIGR00650; MG442; 1.
KW Complete proteome.
SQ SEQUENCE 1189 AA; 137542 MW; 0FGF95636BE31BC8 CRC64;

Query Match 7.5%; Score 114; DB 2; Length 1189;
Best Local Similarity 18.6%; Pred. No. 59;
Matches 62; Conservative 59; Mismatches 130; Indels 82; Gaps 9;

QY 7 EQTEVVK-----SAIETADGALDLYNKYLDQVPWKTFDETIKLSRPFQESQASV 60
Db 163 EESAGVLKYKRAESIQKLDHTEDNLRVEDILYDLEGRVEPLKEAATAKEYKQLSKE 222

QY 61 LVGDIKVLMDSQKYEATQTVVWC-----GVVTTLLSAYILLFDEYNE 106
Db 223 MEQSDVITVSDIDHYTEDNQRLDERLNHLKSQAQEGQAQINQLLQKY----- 273

QY 107 KKASAKDILIRILDGCVKKNLNEAQSLLTSSQSFNNAAGKLLALDSQLTND-----FS 160
Db 274 -KGKROQN-----DYDIEKLN-----YELVKATENEYQSLGNVLEERKKNQSETNARYE 323

QY 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNRL-----K 216
Db 324 EELDNLSEQSDISIKNEKAQNE-----KLLAELKNKQKQLNK 359
```

```
QY 217 TVQNFTLSATVQKANKDIDAALKLATEIAAIGEIKETETTRFRFVVDVDDMLSLK 276
Db 360 EVOELESLLYISDEQDEKLEIKNSYLLMSQSDVNNDIRFLEHTINENAKKSRLDS 419
QY 277 AAKKMINTCNEYQ-----RHGKKTLFEV 300
Db 420 RLVEAFNQLKDIOQNTQTQKEYSQSSKSKMEKV 452

RESULT 36
Q7RD43 PRELIMINARY; PRT; 1611 AA.
AC Q7RD43
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Hypothetical protein.
GN Names=PY05582;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteza M.,
  Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
  Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
  Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
  Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
  Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
  Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
  van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
  Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
  Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
  parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -1- CAUTION: The sequence shown here is derived from an
  EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
  preliminary data.
CC EMBL; AABL01001787; EAA17620.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1611 AA; 188293 MW; CD0493CC4A9BC5D1 CRC64;

Query Match 7.5%; Score 114; DB 2; Length 1611;
Best Local Similarity 21.7%; Pred. No. 85;
Matches 70; Conservative 60; Mismatches 111; Indels 82; Gaps 15;

QY 24 LDLYNKYLDQVIPWKTFDETIK-ELSRFQOEY-----SQEASVLVGDIKVLML----- 70
Db 799 LNIIRNLYLMKK-----NEKIKFENKLLKDYDNLKSKYKNVLPKLCIKFLLKYYTTKD 852

QY 71 -DSQDYFEATQTVVWCGVVVTTLLSAYILLFDEYNEKASAKDILIR-ILDDGVKKLN 128
Db 853 RSYDKKYE-----NMRHLQSEKNLMLIEKKYQEQNNINNMMLSDCLKLQ 903

QY 129 EAQKSLTSSQSFNNAAGKLLALDSQLTNDPSE-KSSYFQSQVD-----RIRKE 176
Db 904 EENESIISNNNSYKTEVETINSKYQHLQNDFNKIKSEHEKLIKIEHKIKRENENIKIEKE 963

QY 177 AYAGAAGIVAGPGLIISYSIAAGVIEG-----KLIPELNRLKTVQNFSTLSATVK 230
Db 964 TLIKELGDTKAKYFNM-----TGILQGBEKMYAKKIKLEBKLOKEKEKKNINIDVK 1016

QY 231 QANKDIDA-AKL-----KLATEIAAIGEIKETETTRFYVD-----YDDLML 271
Db 1017 D-----EIDTFTKILEKENENHKIKELKELNIEEKYENTQINFDSLKKEFEKSFDEVQI 1073

QY 272 SLKGAAGKMWINTCNEYQORHGK 294
Db 1074 IL-----KEMIQEKEYTNSYNK 1091
```

Db 344 IDRLSGKLNTRTRVFLAEAS 366
: ||: : ||
: ||: : ||

RESULT 37

Q9Z429 PRELIMINARY; PRT; 538 AA.
ID Q9Z429
AC Q9Z429
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE MethV1-accepting chemotaxis protein.
GN Name: nahy.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G7;
RX MEDLINE=99255564; PubMed=10322041;
RA Grimm A.C., Harwood C.S.;
RT "Nahy, a catabolic plasmid-encoded receptor required for chemotaxis of
RT Pseudomonas putida to the aromatic hydrocarbon naphthalene.";
RL J. Bacteriol. 181:3310-3316(1999).
DR EMBL; AF100302; AAD13223.1; --
DR HSSP; P02942; 1QUT.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR000727; T_SNARE.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; WA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
DR PROSITE; PS50192; T_SNARE; 1.
DR Plasmid.
KW Complete proteome.
SQ SEQUENCE 538 AA; 58716 MW; C915895E65F45E56 CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 538;
Best Local Similarity 21.9%; Pred. No. 25;
Matches 84; Conservative 56; Mismatches 118; Indels 125; Gaps 20;
QY 9 TVEVVK-SAIETADGALDLYNKVLDQVTPWKTETIKELSRFKQYEQEASVLGDIKV 67
DB 23 TIELMGFSALQR--GVASLNTVYLDVVP-----LRDLKTADLYA-----V 62
QY 68 LMDSQDK-----YFEATQTVVWCVVTVLLSAYILLFDEYNEKASAKQ---DILI 117
DB 63 KIVDSHAKSGRMTVAQAEQVKGAGQIDMLWHAY-----QTKKIDEEQSRVDALA 116
QY 118 RILDDGVKLINEAQKSLTSSQSFNNASGKLALDSQLTND-----PSEKSSYF-QSQ 169
DB 117 KLVDQADPI-ERLKGIL-----ERGDKAALDTFVENMYPLDPLSEGLSHLTQIQ 167
QY 170 VDIRKQAYAGAA-----GIV--AGPFGLLISYSI-----AAGVI 203
DB 168 VEE-SKRAYDAAVVLYDSRTMALLLLGLICGGVFAIRLRSIHPITLTKDAARVA 226
QY 204 EGKLIPEL-----NNRLKTVQNFETSLSATVQKANKDIDAACKLATEIAAIGETET 259
DB 227 LGDLSQSISQVSGRNEVDVQSQVQAMQANRLNTLQDIOGSAQA---AAAELOQTATES 283
QY 260 TR--FYVDYDDLMS-----LLKGAACKWIMTNCNE 287
DB 284 TAQGIHQNDQMOMQAAATVTEMSSAAVDEADNANRTSNASHEAMDLDAGGRKQVMLTRET 343
QY 288 YQRRHGK-----KTLFVEPVDVAS 305

RESULT 38

Q8Z0N2 PRELIMINARY; PRT; 727 AA.
ID Q8Z0N2
AC Q8Z0N2
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE All0059 protein.
GN OrderedLocusNames=all0059;
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003581; BAB77583.1; --
DR PIR; AC1814; AC1814.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003856; LPS_Wzz_MPA.
DR Pfam; PF02706; Wzz; 1.
KW Complete proteome.
SQ SEQUENCE 727 AA; 82555 MW; B939994D1F4656D8 CRC64;

RESULT 39

Q9YVT6 PRELIMINARY; PRT; 1127 AA.
ID Q9YVT6
AC Q9YVT6
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein MSV156.

Wed Jan 5 14:16:59 2005

us-09-993-292b-2.rup

```
QY 193 IISYSIAAGVIEGKLIPELNNRLKTVQNFSTL-----SATVQKANKD 235
Db 436 -IS-NLSIDFQETKL-KENGDKLSTLKLLESLNQKNQNIERDKKELDEVNSKIQBIERD 492
QY 236 IDAAK-----IKLATETAAGEI-KTETETTFYVDYDMLSLKGAACKMINTCNEYQQ 290
Db 493 VNOSKKNYEIGIVEKINEIAEANKRIESTK-----ELIQTIQNLISTQBELLK 542
QY 291 --RHGKK 295
Db 543 NIEGNK 549
RESULT 42
ID Q6C359 PRELIMINARY; PRT; 1906 AA.
AC Q6C359;
DT 01-OCT-2004 (TREMREL. 28, Created)
DT 01-OCT-2004 (TREMREL. 28, Last sequence update)
DT 01-OCT-2004 (TREMREL. 28, Last annotation update)
DE Similar to DEHAOC096589 Debaryomyces hansenii IPF 1836.1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisarame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RN Nature 430:35-44(2004).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382132; CAG77705.1; -.
SQ SEQUENCE 1906 AA; 210346 MW; F6ED7D1AF7B7562B CRC64;
Query Match 7.5%; Score 113.5; DB 2; Length 1906;
Best Local Similarity 22.1%; Pred. No. 1.1e+02;
Matches 66; Conservative 48; Mismatches 116; Indels 69; Gaps 12;
QY 7 EQTVEVKSIAETADGALDLYNKV-----DOVIPMTKTFDTIKELSRFQEQYSQBSAIVL 62
Db 1195 EKELETKSTLETKTAELTKSKELTAKSDEA---TTSYAKVKELTSSAALEKKQTLLK 1251
QY 63 GDIKVLMDSDQKYFEATQTVYEMCVGVVQTLLSAYILLDFEYNEKKAQAKQILYILD 122
Db 1252 AMADNLTKDLAEK-----TKELVAKSEL---ESSNTSKEEVDVLTKKLSD 1295
QY 123 GVKKLNAQKSLTSSQFNASKLLALDSQLTNDFSEKSSVFQSOVDRIKAEVAGAA 182
Db 1296 ATAFAVELKKS---SQAAETASKVSALAEKUTK-ASESS---KAECLKVNK----- 1341
QY 183 AGIVAGPFPGLIISYSIAAGVIE-----GKLIPELNNRLKTVQNF---FTLSATVQKA 232
Db 1342 -----LLSSPFKEKLTQSKDHDHSTEVSKLTQEVRESTLKAENFEHDTISLKDQLAQA 1392
```

```
QY 233 NKDIDAAKLKLAIEAIGIKTETETTFYVDYDMLSLKGAACKMINTCNEYQOR 291
Db 1393 EKERDALRELTDSIK-----EMENERTSLTKDAD-----SATKELTNKVSMLQTK 1438
RESULT 43
ID O45614 PRELIMINARY; PRT; 3102 AA.
AC O45614; P91824; Q9TZR4; 06, Created)
DT 01-JUN-1998 (TREMREL. 22, Last sequence update)
DT 01-OCT-2002 (TREMREL. 28, Last annotation update)
DT 01-OCT-2004 (TREMREL. 28, Last annotation update)
DE C. elegans LAM-3 protein (Corresponding sequence T22A3.8) (Laminin
DE alpha chain) (Hypothetical protein T22A3.8).
GN Name=lam-3; Synonyms=lama1/2; ORFNames=T22A3.8;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RN Science 282:2012-2018(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Barlow K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=22668894; PubMed=12783803;
RA Huang C.C., Hall D.H., Hedgecock E.M., Kao G., Karantz V.,
RA Vogel B.S., Hutter H., Chisholm A.D., Yurchenco P.D., Wadsworth W.G.;
RA "Laminin alpha subunits and their role in C. elegans development.";
RL Development 130:3343-3358(2003).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.A.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL008585; CAA15432.3; -.
DR EMBL; Z81125; CAA15432.3; JOINED.
DR EMBL; AL008585; CAB03385.3; JOINED.
DR EMBL; Z81125; CAB03385.3; -.
DR EMBL; AF074902; AAC26793.1; -.
DR PIR; F87908; F87908.
DR PIR; T23064; T23064.
DR PIR; T43291; T43291.
DR HSSP; P02468; INPE.
DR WormPep; T22A3.8; CE31067.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0005606; C:laminin-1; IEA.
DR GO; GO:0005102; F:receptor binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030155; P:regulation of cell adhesion; IEA.
DR GO; GO:0030334; P:regulation of cell migration; IEA.
DR GO; GO:0045995; P:regulation of embryonic development; IEA.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR009254; Laminin_I.
DR InterPro; IPR010307; Laminin_II.
DR InterPro; IPR008211; Laminin_N.
```

```

DR InterPro: IPR003129; TSP_N.
DR Pfam: PF00052; Laminin_B; 2.
DR Pfam: PF00053; Laminin_EGF; 17.
DR Pfam: PF02210; Laminin_G; 3.
DR Pfam: PF06008; Laminin_I; 1.
DR Pfam: PF06009; Laminin_II; 1.
DR Pfam: PF06009; Laminin_III; 1.
DR Pfam: PF00055; Laminin_N; 1.
DR PRINTS: PR00011; EGFLAMININ.
DR ProDom: PD003031; Laminin_B; 2.
DR SMART: SM00180; EGF_Lam; 17.
DR SMART: SM00281; LamB; 2.
DR SMART: SM00282; LamG; 4.
DR SMART: SM00136; LamNT; 1.
DR PROSITE: PS00022; EGF_1; 12.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 14.
DR PROSITE: PS50025; LAM_G_DOMAIN; 3.
DR Hypothetical protein; Laminin EGF-like domain.
SQ SEQUENCE 3102 AA; 343790 MW; 2E9305D5C591FAB CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 3102;
Best Local Similarity 20.8%; Pred. No. 2e+02;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

QY 25 DLYNKYLDQVLPW-KTFDETIKELSRFKQYSQASVLVG-----DIKVLIMDS 72
DB 1881 DLKGR-IDVLEQWMDYRETIYDVSCKDTADAERMSLVGGRINRYKEVSEIEKURVEA 1939

QY 73 QDK-----YFEATQTV-----YEWCVVVTQLLSAYILLFDEY 104
DB 1940 EDQIAYSENSIEKARSELNMWFEDEKINWTLAELPDVLEQCNITLL---YSOLIDEY 1996

QY 105 NEK-----KASAQKDLIRILD-----DGVKVLNEAQS 133
DB 1997 DEEYVQTAGRHAELKLEVQAQK-IVDFEVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055

QY 134 LTTSSQSFNNASGKLALDSQLTNDFSEKSYFQSOVDRIKAYAGAAAGIVAGPGLI 193
DB 2056 AAASE-----AVSKMLSEGESGDANES--LRSQLKLNESLSN----- 2097

QY 194 ISYSIAAGVIE--GKLPENLRKLTQNFFTLSATVQKQKIDDAKLLATEIAAIG 251
DB 2098 VDSNNAKIVIELKKEKKDLTDLGLHNLKTSI-----VKRLG 2136

QY 252 EIKTETETTPYVDYDLMISLKGAAKQM-INTCNEYQQRHGKT 296
DB 2137 VIKNEASS--WDDKHDRMHSILKNGAKTAHRSANVKESEGIKT 2179

RESULT 44
Q73AJ3 PRELIMINARY; PRT; 465 AA.
AC Q73AJ3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Flagellin.
GN OrderedLocusNames=BCE1780;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Raiko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L., Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis px01.";
RL Nucleic Acids Res. 32:977-988 (2004).
DR ENBL; AE017269; AAS40708.1; --
DR TIGR; BCE1780; --
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

Query Match 7.5%; Score 113; DB 2; Length 465;
Best Local Similarity 22.0%; Pred. No. 23;
Matches 72; Conservative 53; Mismatches 116; Indels 86; Gaps 15;

QY 6 AEQTVVVKSAIETADGALDLYNKYLDQV---IPWKTFTDETIKELSRFK-----Q 52
DB 184 AKATI-TAKDAFDAAKDASDAKKEIDAAAKOTPSKNDAAQLAKEYTEAKATLATLKPTDA 242

QY 53 EYSQASVL-----VGDIKVLIMDSQDKYFEATQTVYEWCVVVTQLLSAYILLFDEYNE 106
DB 243 TYAAKAAELDAATTAALNDNAKVLVDGYEKKLTTTKTEA-----EYTA 285

QY 107 KKAQAQKDLIRILDGKVKLNEAQSLLTSSQSF-NNASGKLLA---LSQLT-----N 157
DB 286 AKQSTK-----STAAADLVTKYETAKSNALGNDIAKEYLEAKTAYEANKN 331

QY 158 DFSEKSYFQSOVDRIKAYAGAAAGIVAGPGLIISYSIA--AGVIECKL--IPELNN 213
DB 332 DISSKR-FAAATELNKOITANKAAKV-----LVETYEKAKTAGTETKSLVAVDKIDE 384

QY 214 RLKTVQNFFTLSATVQKQKIDDAKLLATEIAAIGEI-----KTETETTRFYVDYD 267
DB 385 ALKTIADNRATLGATLNRLDFVNNLKSQASASASQIEDADMAKEMSEMTKF----- 439

QY 268 DMLSLKGAAKQMINTCNEYQQRHGK 294
DB 440 ---KILNEAGISMLSQANQTPQWVSK 462

RESULT 45
AAS40708 PRELIMINARY; PRT; 465 AA.
AC AAS40708;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Flagellin.
GN BCE1780.
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Raiko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L., Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis px01.";
RL Nucleic Acids Res. 32:977-988 (2004).
DR ENBL; AE017269; AAS40708.1; --
DR TIGR; BCE1780; --
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

Query Match 7.5%; Score 113; DB 2; Length 465;
Best Local Similarity 22.0%; Pred. No. 23;
Matches 72; Conservative 53; Mismatches 116; Indels 86; Gaps 15;

QY 6 AEQTVVVKSAIETADGALDLYNKYLDQV---IPWKTFTDETIKELSRFK-----Q 52
DB 184 AKATI-TAKDAFDAAKDASDAKKEIDAAAKOTPSKNDAAQLAKEYTEAKATLATLKPTDA 242

QY 53 EYSQASVL-----VGDIKVLIMDSQDKYFEATQTVYEWCVVVTQLLSAYILLFDEYNE 106
DB 243 TYAAKAAELDAATTAALNDNAKVLVDGYEKKLTTTKTEA-----EYTA 285

InterPro: IPR001492; Flagellin_N.
DR Pfam: PF00700; Flagellin_C; 1.
DR Pfam: PF00669; Flagellin_N; 1.
DR PRINTS: PR00207; FLAGELLIN.
DR ProDom: PD000316; Flagellin_C; 1.
DR Complete proteome; Flagellum.
SQ SEQUENCE 465 AA; 49855 MW; E4C307E7717381D5 CRC64;

```

Query Match 7.5%; Score 113; DB 1; Length 956;
 Best Local Similarity 24.4%; Pred. No. 53;
 Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;

QY 107 KKSAQKDIILIRILDDGVKKLNEAKSLLTSSQSF--NNASGKLLA---LDSQLT-----N 157
 DB 286 AKQSTK-----STRAADLVTKYETAKSNALGNDIAKEYLEAKTAYEANKN 331
 QY 158 DFGEKSSYFQSDVRIRKEAYAGAAAGIVAGPPGLLIISYSIA--AGVIEGKL--IPELNN 213
 DB 332 DISSKSR--FEAAETELNKDITANKAAKV-----LVETYEKAKTAGTTEKSLVAVDKIDE 384
 QY 214 RLKTVQNFPSLSNTVKQAKNDIDAAKLIKLATBIAAIGEI-----KTETETTRFRVVDYD 267
 DB 385 ALKTIADNRATLGATLNRDLFNVNVLKASQASMASASQIEDADMAKENSEWTKP-----439
 QY 268 DLMELSLKGAAKKMTNCNEYQQRHGK 294
 DB 440 ----KILNEAGISLMSQANQTPQVMSK 462

RESULT 46
 YEF3 YEAST
 ID YEF3 YEAST STANDARD; PRT; 956 AA.
 AC P32618;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DE Hypothetical 106.1 kDa protein in GLY1-GDAI intergenic region.
 GN OrderedLocNames=YEL043W; ORFNames=SYGP-ORF14;
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=9713264; PubMed=9169868;
 RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
 RA Araujo R., Aviles E., Berto A., Brennan T., Carpenter J., Chen E.,
 RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
 RA Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
 RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
 RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S.,
 RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome V";
 RT Nature 387:78-81(1997).
 RN [2]
 RN FIBRONECTIN TYPE III DOMAIN
 RP MEDLINE=97148176; PubMed=8994808;
 RA Bateman A., Chothia C.;
 RA "Fibronectin type III domains in yeast detected by a hidden Markov
 model";
 RT Curr. Biol. 6:1544-1546(1996).
 RL CC
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U18779; AAB64999.1; --
 CC PIR; S30834; S30834.
 CC
 CC SCD; S0000769; YEL043W.
 CC GO; GO:0005783; C:endoplasmic reticulum; IDA.
 CC InterPro; IPR003961; FN III.
 CC InterPro; IPR008957; FN_III-like.
 CC SMART; SM00060; FN3; 1.
 KW Hypothetical protein.
 FT DOMAIN 35 125
 FT SEQUENCE 956 AA; 106132 MW; 3F78B09A0FCA03AF CRC64;

Query Match 7.5%; Score 113; DB 1; Length 956;
 Best Local Similarity 24.4%; Pred. No. 53;
 Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;

QY 38 KTDFTIKELSRFKQYSQEAASVLDGDIKVLMD-----SQDKYFEATQTVYEWGCVVQ 92
 DB 272 KSLKSNIKSL-----ENSKLLTDLKEKLNKIDKSKKISKORNDQMKSQEDTE 322
 QY 93 LLS-----AYILLFDEYNEKKAQAQDI-----LIRILDDGVKKLNEAKSLLTSS--S 138
 DB 323 LLSKDTIKKFKYKLLNESNASVANINKEIESLQNEISKMEESNKLNASKSLITSIVN 382
 QY 139 QSFNN-----ASGKLLALDSQITNDFS--EKSSYF-----QSOVDRIKAEYA 179
 DB 383 ANVENDKPIASGELSALVKKL--NDFLEKNGFLSNAGEEFLSKLNADSSLIKMKOE---438
 QY 180 GAAAGIVAGPPGLLIISYSIAA-----GVIEGKLIP--ELNNR--LKT-----V 218
 DB 439 -----LSIDQLEANWKLQRSNLLKKLSALENQFNWNLNRLNLTLMVQPY 486

RESULT 47
 RA50 METJA STANDARD; PRT; 1005 AA.
 ID RA50 METJA
 AC Q58718;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE DNA double-strand break repair rad50 ATPase.
 GN Name=rad50; OrderedLocNames=MJ1322;
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Adams M.D., Reich C.I.,
 RA Karlavagne A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Glodek A.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
 RA Usterback T.R., Kelley J.M., Peterson J.D., Kaine B.P., Borodovsky M.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RT Science 273:1058-1073(1996).
 RL CC
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
 CC rad50/mre11 complex possesses single-strand endonuclease activity
 CC and ATP-dependent double-strand-specific 3'-5' exonuclease
 CC activity. Rad50 provides an ATP-dependent control of mre11 by
 CC unwinding and/or repositioning DNA ends into the mre11 active site
 CC (By similarity).
 CC
 CC -!- COPACTOR: Binds 1 zinc ion per heterotetramer (By similarity).
 CC
 CC -!- SUBUNIT: Heterotetramer composed of two mre11 subunits and two
 CC rad50 subunits (By similarity).
 CC
 CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U67572; AAB99331.1; --
 CC
 CC DR

```

DR PTR: A64465; A64465.
DR HSP: P58301; 1F2T.
DR TIGR: MJ1322; -.
DR HAMAP: MF_00449; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR007517; Rad50_zn_hook.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF04423; Rad50_zn_hook; 1.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRFAMs: TIGR01725; phage_HK97_gp10; 1.
KW ATP-binding; Coiled coil; Complete proteome; DNA repair; Hydrolase;
KW Zinc.
FT NP_BIND 32 39 ATP (By similarity).
FT DOMAIN 158 849 Coiled coil (Potential).
FT METAL 502 502 Zinc (By similarity).
FT METAL 505 505 Zinc (By similarity).
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 7.5%; Score 113; DB 1; Length 1005;
Best Local Similarity 19.7%; Pred. No. 57;
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;

QY 7 EQTEVVVKSATETADGALDLYNKYLDQVTPKTFDE-----TIKELSRF-----KQEYSQ 56
DB 605 DEILEDIKSOLNPKF---NFNQYLSAVSYLVNSVDEEGIRNRIKEIENIVSGWNKCKRE 661

QY 57 EASVL-----VGDIKVLMDSQKFEATQTVYEWGVVTVLLSAVILLPDEYNEKKAS 110
DB 662 ELNKLREDEIREINKDLKLNKKE-----LIEIENRRSLKFPDKYKYLGL 710

QY 111 AQDILIRILDGKVLNKAQKSLTSSQSFNNASGKLLALD----- 152
DB 711 TEKLEEKLNKIDGLEET-----YNICNSKILADINIRKYNKEDIYIYNLK 757

QY 153 -----SQTNDPSEKSSYFQSQVDRIKREAYAGAAAGIVAGPFGLIISYIAAGVIEGKLI 208
DB 758 ILEVNKEINDEIERISYINQKLD-----INNEE-----EHKXI 792

QY 209 PEL--NRR--LKTQVNFVTSATVKAQKIDIAKULATETAIGETITETTFYV 264
DB 793 KELYENKQBLDNVREQKTEITGIEYLLKQVESLKRL-----KEMSNLEKEKEKLTQFV 848

QY 265 DYDD 268
DB 849 EYLD 852

RESULT 48
NAGH_CLOPE STANDARD; PRT; 1628 AA.
AC P26831;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase) (Mu toxin).
GN Name=nagh; OrderedLocusNames=CPE0191;
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=94232189; PubMed=8177218;
RA Canard B., Garnier T., Saint-Joanis B., Cole S.T.;
RT "Molecular genetic analysis of the nagH gene encoding a hyaluronidase of Clostridium perfringens.";
```

```

RL Mol. Gen. Genet. 243:215-224 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=13 / Type A;
RC PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayaishi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
CC -1- FUNCTION: Putative virulence factor which is likely to act on connective tissue during gas gangrene.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1->4-linkages between N-acetyl-beta-D-glucosamine and D-glucuronate residues in hyaluronate.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Contains 1 F5/8 type C domain.
CC -1- CAUTION: The partially purified protein from strain CPN50 is approximately 70 kDa smaller than the sequence indicated here.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81878; AAA23259.1; -.
DR EMBL: AP003185; BAB79897.1; -.
DR PIR: S43904; S43904.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR000421; FAS8_C.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF00754; F5_F8 type C; 3.
DR SMART: SM00231; F58C_1; -.
DR PROSITE: PS00022; FAS8C_3; 1.
KW Complete proteome; Glycosidase; Hydrolase; Signal; Toxin; Virulence.
FT SIGNAL 1 30 potential.
FT CHAIN 31 1628 Hyaluronoglucosaminidase.
FT DOMAIN 781 953 F5/8 type C.
FT VARIANT 147 147 G -> A (in strain CPN50).
FT VARIANT 172 175 KIQS -> EIKN (in strain CPN50).
FT VARIANT 250 250 V -> M (in strain CPN50).
FT VARIANT 548 548 A -> E (in strain CPN50).
FT VARIANT 558 558 D -> E (in strain CPN50).
FT VARIANT 614 614 G -> S (in strain CPN50).
FT VARIANT 944 944 I -> V (in strain CPN50).
FT VARIANT 950 950 N -> S (in strain CPN50).
FT VARIANT 979 979 T -> I (in strain CPN50).
FT VARIANT 982 982 I -> L (in strain CPN50).
FT VARIANT 1042 1042 I -> F (in strain CPN50).
FT VARIANT 1043 1628 Missing (in strain CPN50).
SQ SEQUENCE 1628 AA; 182474 MW; D4252A2512BBED69 CRC64;

Query Match 7.5%; Score 113; DB 1; Length 1628;
Best Local Similarity 23.2%; Pred. No. 1e+02;
Matches 72; Conservative 42; Mismatches 112; Indels 84; Gaps 14;

QY 27 YNKYLQVTPKTFDETIKELSRFKQEYSQAS-VLVGDIKVLMDSQKFEATQTVYE 85
DB 1328 YNNIYEIKIEWTVAPNIYEIITLNOEFEPFVNDLAKYDELINUSGDEY----- 1378
QY 86 WCGVTVQLLSAYILLDFEYNEKKA-----SAQKDILIRILDGKVLNKAQKSL-----LT 136
DB 1379 -----TSSFTLKEALNEAKSLDDSSQKXI-----DKALEKLNKAERELDLRAT 1426
QY 137 SSQSFNNAAGKLLALDSQLTND-----FSE-----KESYFQSQVDRIKRE 176
DB 1427 DFEDEFN---KVLTLGNSLVEEYTAESWALFSEAAANEAKNKADYTDQIQINQIVID 1482
QY 177 AYAGAAGIVAGP-----FGLIISY--STAAGVIEGKLIPE-----LNNELKTVQN 220
```

Db 1483 LDASIKALVETPEVDKTNLGINOQKSLLDSEVGFVGHKGAKDGLTVEINKAE 1542

Qy 221 FFTSLSATVQKANDDAKLKLAIEAATGEIKTETTRF-----YVDYDMLSLK 276

Db 1543 VFNKEDAT---BEEINLAKESLEGAIARFNSLLIBESTGDFNGNGKIDIGDLAM----- 1593

Qy 277 AAKMINTCN 286

Db 1594 VSKNIGSTN 1603

RESULT 49

Q71X69 PRELIMINARY; PRT; 927 AA.

AC Q71X69;

DT 05-JUL-2004 (TREMELrel. 27, Created)

DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)

DE Membrane protein, putative.

GN OrderedlocusNames=LMOF2365.2330;

OS Listeria monocytogenes (serotype 4b / strain F2365).

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=265669;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=15115801; DOI=10.1093/nar/gkh562;

RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,

RA Kolonay J.F., Raeko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,

RA Peterson J.D., White O., Nelson W.C., Niemman W.C., Beanan M.J.,

RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,

RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,

RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Uhlich G.A.,

RA Bayles D.O., Luchansky J.B., Fraser C.M.;

RA "Whole genome comparisons of serotype 4b and 1/2a strains of the food-

RT borne pathogen Listeria monocytogenes reveal new insights into the

RT core genome components of this species.";

RL Nucleic Acids Res. 32:2386-2395(2004).

DR EMBL; AE017330; AAT05096.1; -

DR InterPro; IPR004089; Chmtaxis transd.

DR Pfam; PF00015; MCPsignal; 1.

DR ProDom; PD001127; Outsurface; 1.

KW Complete proteome.

SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;

Query Match 7.4%; Score 112.5; DB 2; Length 927;

Best Local Similarity 20.7%; Pred. No. 56;

Matches 78; Conservative 43; Mismatches 135; Indels 121; Gaps 14;

Qy 1 MTSIFAQTVVVKSAIETADGALDLYNKYLQVVPW----KTFDETIKELSRFKQEYSQ 56

Db 160 VTKSYAEAFDKIKESGEGFAAGDSGKIKDGLVKSQEGNKTIISTNLKTLADSSLTFKD 219

Qy 57 EASVL-----VGDIK-----VLLMDSQDK 75

Db 220 GANTLEVGLKTYTDGVTNTAAAGDKLNDGVSTLAAGVGLKGVAAALDGGATKLSGVST 279

Qy 76 YFEATQTVVWCWVVTQLLSAYILLFDEYNEKASAKDILIRILDGQVKLNEAKSLL 135

Db 280 YTSQVDTL---AGINQAVTGSTALSDGLNKNGS-----VPTLASGITQLNNGOKSLA 330

Qy 136 TSSQSFNNAGSKLLA---LDSQLTN-----DFSEKSSYFQSQVDRIKEAYAGAAAGI 185

Db 331 TGLDSLVDGSKLSAGLKELDGNLTDKQKIAQLKQGMNDLQGGIDQLNKSNGVEDAA-- 388

Qy 186 VAGPFGLLISYSTAAGVIEGKLIPELNRL-----KTVQNFFTSLSATVKQ 231

Db 389 -----LAKQLAA---LQKSLDGLQNGLTFFIKSNANFDAEAIKSKINATAGVSAEDKQ 437

Qy 232 -----ANKQIDAAKLKLATEIAAIGEIKT-----ETETTRFYVDYDMLSLK 276

Db 438 KIIDAQADLDKETQKSATQVATVEQLQSGLSGLDLAAIQTQVT-----ELQTG 486

Qy 277 AAKMINTCNEYQQRHG 293

Db 487 VAKISAG-----YQAVHG 499

RESULT 50

AAT05096 PRELIMINARY; PRT; 927 AA.

AC AAT05096;

DT 10-MAY-2004 (TREMELrel. 27, Created)

DT 10-MAY-2004 (TREMELrel. 27, Last sequence update)

DT 10-MAY-2004 (TREMELrel. 27, Last annotation update)

DE Membrane protein, putative.

GN LMOF2365.2330.

OS Listeria monocytogenes str. 4b F2365.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria;

OC Listeria monocytogenes.

OX NCBI_TaxID=265669;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=4b F2365;

RX PubMed=15115801;

RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,

RA Kolonay J.F., Raeko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,

RA Peterson J.F., White O., Nelson W.C., Niemman W.C., Beanan M.J.,

RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,

RA Haft D.H., Selengut J., Van Aken S., Khouri H., Fedorova N.,

RA Forberger H., Tran B., Kathariou S., Wonderling L.D., Uhlich G.A.,

RA Bayles D.O., Luchansky J.B., Fraser C.M.;

RA "Whole genome comparisons of serotype 4b and 1/2a strains of the food-

RT borne pathogen Listeria monocytogenes reveal new insights into the

RT core genome components of this species.";

RL Nucleic Acids Res. 32:2386-2395(2004).

DR EMBL; AE017330; AAT05096.1; -

SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;

Query Match 7.4%; Score 112.5; DB 2; Length 927;

Best Local Similarity 20.7%; Pred. No. 56;

Matches 78; Conservative 43; Mismatches 135; Indels 121; Gaps 14;

Qy 1 MTSIFAQTVVVKSAIETADGALDLYNKYLQVVPW----KTFDETIKELSRFKQEYSQ 56

Db 160 VTKSYAEAFDKIKESGEGFAAGDSGKIKDGLVKSQEGNKTIISTNLKTLADSSLTFKD 219

Qy 57 EASVL-----VGDIK-----VLLMDSQDK 75

Db 220 GANTLEVGLKTYTDGVTNTAAAGDKLNDGVSTLAAGVGLKGVAAALDGGATKLSGVST 279

Qy 76 YFEATQTVVWCWVVTQLLSAYILLFDEYNEKASAKDILIRILDGQVKLNEAKSLL 135

Db 280 YTSQVDTL---AGINQAVTGSTALSDGLNKNGS-----VPTLASGITQLNNGOKSLA 330

Qy 136 TSSQSFNNAGSKLLA---LDSQLTN-----DFSEKSSYFQSQVDRIKEAYAGAAAGI 185

Db 331 TGLDSLVDGSKLSAGLKELDGNLTDKQKIAQLKQGMNDLQGGIDQLNKSNGVEDAA-- 388

Qy 186 VAGPFGLLISYSTAAGVIEGKLIPELNRL-----KTVQNFFTSLSATVKQ 231

Db 389 -----LAKQLAA---LQKSLDGLQNGLTFFIKSNANFDAEAIKSKINATAGVSAEDKQ 437

Qy 232 -----ANKQIDAAKLKLATEIAAIGEIKT-----ETETTRFYVDYDMLSLK 276

Db 438 KIIDAQADLDKETQKSATQVATVEQLQSGLSGLDLAAIQTQVT-----ELQTG 486

Qy 277 AAKMINTCNEYQQRHG 293

Db 487 VAKISAG-----YQAVHG 499

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:58:34 ; Search time 62.2944 Seconds
(without alignments)
1749.716 Million cell updates/sec

Title: US-09-993-292B-28
Perfect score: 1508
Sequence: 1 MTEIVADKTVVVKNAIETA.....TCNEYQKHGKTLFVPEV 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

- Database :
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1402	93.0	305	9	US-09-993-292A-2
2	126	8.4	1095	14	US-10-128-714-3039
3	126	8.4	1277	14	US-10-128-714-8039
4	120.5	8.0	891	14	US-10-369-493-10918
5	120.5	8.0	962	13	US-10-087-192-222
6	119	7.9	889	16	US-10-437-963-197045
7	118	7.8	961	14	US-10-080-608A-66
8	118	7.8	961	14	US-10-370-688-155
9	117.5	7.8	496	16	US-10-767-701-43815
10	114.5	7.6	1189	15	US-10-282-122A-70920
11	113	7.5	458	16	US-10-467-421-37
12	112.5	7.5	1020	15	US-10-282-122A-12
13	112.5	7.5	1679	14	US-10-369-493-22080

14	112	7.4	1875	14	US-10-369-493-22285	Sequence 22285, A
15	111.5	7.4	2823	14	US-10-369-493-5220	Sequence 5220, Ap
16	111.5	7.4	2823	14	US-10-369-493-5221	Sequence 5221, Ap
17	111	7.4	435	10	US-09-978-309A-80	Sequence 80, Appl
18	111	7.4	476	10	US-09-978-309A-77	Sequence 77, Appl
19	111	7.4	794	15	US-10-257-377-2	Sequence 2, Appli
20	111	7.4	938	15	US-10-282-122A-71606	Sequence 71606, A
21	111	7.4	1023	10	US-09-884-696-5	Sequence 5, Appli
22	111	7.4	1171	14	US-10-032-585-7519	Sequence 7519, Ap
23	110	7.3	1163	15	US-10-282-122A-51864	Sequence 51864, A
24	109.5	7.3	420	15	US-10-425-114-60225	Sequence 60225, A
25	109.5	7.3	660	10	US-09-841-260-139	Sequence 139, App
26	109.5	7.3	660	13	US-10-007-693-139	Sequence 139, App
27	109.5	7.3	660	16	US-10-762-058-139	Sequence 139, App
28	109	7.2	1005	14	US-10-369-493-1061	Sequence 1061, Ap
29	108.5	7.2	476	10	US-09-978-309A-79	Sequence 79, Appl
30	108.5	7.2	552	17	US-10-739-930-6691	Sequence 6691, Ap
31	108.5	7.2	647	15	US-10-425-114-51665	Sequence 51665, A
32	108.5	7.2	886	14	US-10-369-493-1016	Sequence 1016, Ap
33	108	7.2	737	16	US-10-437-963-107270	Sequence 107270, A
34	107.5	7.1	630	15	US-10-205-647A-6	Sequence 6, Appli
35	107.5	7.1	631	10	US-09-978-309A-48	Sequence 48, Appl
36	106.5	7.1	680	15	US-10-425-114-68086	Sequence 68086, A
37	106.5	7.1	1999	14	US-10-028-249A-107	Sequence 107, App
38	106.5	7.1	1999	15	US-10-107-782-107	Sequence 107, App
39	106	7.0	1583	16	US-10-408-765A-1635	Sequence 1635, Ap
40	106	7.0	1961	14	US-10-028-248A-105	Sequence 105, App
41	106	7.0	1961	15	US-10-107-782-105	Sequence 105, App
42	106	7.0	2076	9	US-09-815-242-5815	Sequence 5815, Ap
43	106	7.0	2086	9	US-09-815-242-5639	Sequence 5639, Ap
44	106	7.0	2186	9	US-09-815-242-12913	Sequence 12913, A
45	106	7.0	5795	9	US-09-815-242-12610	Sequence 12610, A
46	105.5	7.0	459	15	US-10-424-599-248508	Sequence 248508, A
47	105.5	7.0	742	15	US-10-282-122A-53523	Sequence 53523, A
48	105.5	7.0	815	16	US-10-437-963-191043	Sequence 191043, A
49	105.5	7.0	873	14	US-10-369-493-11332	Sequence 11332, A
50	105.5	7.0	1312	14	US-10-393-602-148	Sequence 148, App

ALIGNMENTS

RESULT 1
US-09-993-292A-2
; Sequence 2, Application US/0993292A
; Patent No. US20020146430A1
; GENERAL INFORMATION:
; APPLICANT: James B. Galen
; APPLICANT: University of Maryland
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF PROTEINS
; FILE REFERENCE: UOFMD.007A
; CURRENT APPLICATION NUMBER: US/09/993,292A
; CURRENT FILING DATE: 2001-11-23
; PRIOR APPLICATION NUMBER: 60/252,516
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Salmonella Typhi
US-09-993-292A-2

Query Match 93.0%; Score 1402; DB 9; Length 305;
Best Local Similarity 91.1%; Pred. No. 4.5e-107;
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVIPHQTFDETIKLSRFKQYSQASV 60
Db 1 MTSIFAQTVVVKNAIETADGALDLYNKYLDQVIPHQTFDETIKLSRFKQYSQASV 60
Qy 61 LVGDIKTLMDSQDKFEATQTVVWCVATQTLAAVILLFDEYNEKASQAQDILIKVL 120

Db 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQQLSAYILLDFDYNKSKASQAQDILIRIL 120
Qy 121 DDGITKLNEAQSLLVSSQSFNNASGKLLALDQLNDSEKSSYFQSQVDKIRKEAYAG 180
Db 121 DDGIVKLNKAEQSLTSSQSFNNASGKLLALDQLNDSEKSSYFQSQVDKIRKEAYAG 180
Qy 181 AAGVWAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVVKQANKDIDAAK 240
Db 181 AAGIVAGPGLIISYSIAAGVVEGKLIPELNNRLKTQVQNFFTSLSATVVKQANKDIDAAK 240
Qy 241 LKLTETIAAIGEIKTETETTRFYVDYDMLSLILKEAAKWMINTCNEYQKRHGKTLFEV 300
Db 241 LKLTETIAAIGEIKTETETTRFYVDYDMLSLILKEAAKWMINTCNEYQKRHGKTLFEV 300
Qy 301 PEV 303
Db 301 PDV 303

RESULT 2
US-10-128-714-3039
; Sequence 3039, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3039
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3039

Query Match 8.4%; Score 126; DB 14; Length 1095;
Best Local Similarity 21.4%; Pred. No. 0.26; Mismatches 73; Indels 146; Gaps 15;
Matches 73; Conservative 49;
Qy 4 IVADKTVVVKNAIETADGALDLYNKVLDQVWPQTFDETIKELSRFKQYEQASV--L 61
Db 281 IKKEKDIEEATNALVPVDEKVDITRKVER-----PASRIABIG--KERDSQAANVKQL 332
Qy 62 VGDITKLMDSDQKYFEATQTVVWCGVATQLLAAVILLDFEYNEKASQAQDILIKVLD 121
Db 333 EKDLK--VVEKAQAQWEA-----EW----- 350
Qy 122 DGITKLNEAQSLLVSSQSFNNASGKLLALDQ-----LTNDFSEKSSYFQSQVDKIRKEA 177
Db 351 -----HKTMSNKGQQLSESDQEQYKMLKEVSKRSSAEQINLDNLKRQR 394
Qy 178 YAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNFFT----- 224
Db 395 KTEAEA-----YN-----SLKSFDFSTEWQKSVENDTQTLTERKS 430

Qy 225 -LSNTVKQANKDIDAAK---LTTEIAAIGEIKTETETTRFYVDYDMLSLILKEA--- 277
Db 431 ALNDVTKTSKEIDRKKELNALTSLRISRQMRTELE-----EKQVVVLKKLEADDG 484
Qy 278 -----AKWMINT-----CNEYQKRHGK 294
Db 485 KKQTERELRAKELISTLKRIFPGVGRVSDLCRPFQKKYAE 525
RESULT 3
US-10-128-714-8039
; Sequence 8039, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8039
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8039
Query Match 8.4%; Score 126; DB 14; Length 1277;
Best Local Similarity 21.4%; Pred. No. 0.32; Mismatches 73; Indels 146; Gaps 15;
Matches 73; Conservative 49;
Qy 4 IVADKTVVVKNAIETADGALDLYNKVLDQVWPQTFDETIKELSRFKQYEQASV--L 61
Db 323 IKKEKDIEEATNALVPVDEKVDITRKVER-----PASRIABIG--KERDSQAANVKQL 374
Qy 62 VGDITKLMDSDQKYFEATQTVVWCGVATQLLAAVILLDFEYNEKASQAQDILIKVLD 121
Db 375 EKDLK--VVEKAQAQWEA-----EW----- 392
Qy 122 DGITKLNEAQSLLVSSQSFNNASGKLLALDQ-----LTNDFSEKSSYFQSQVDKIRKEA 177
Db 393 -----HKTMSNKGQQLSESDQEQYKMLKEVSKRSSAEQINLDNLKRQR 436
Qy 178 YAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNFFT----- 224
Db 437 KTEAEA-----YN-----SLKSFDFSTEWQKSVENDTQTLTERKS 472
Qy 225 -LSNTVKQANKDIDAAK---LTTEIAAIGEIKTETETTRFYVDYDMLSLILKEA--- 277
Db 473 ALNDVTKTSKEIDRKKELNALTSLRISRQMRTELE-----EKQVVVLKKLEADDG 526
Qy 278 -----AKWMINT-----CNEYQKRHGK 294
Db 527 KKQTERELRAKELISTLKRIFPGVGRVSDLCRPFQKKYAE 567

Wed Jan 5 14:16:57 2005

```

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 155
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-370-685-155

Query Match      7.8%; Score 118; DB 14; Length 961;
Best Local Similarity 22.9%; Pred. No. 1;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;

Db 426 ESNEQFASQVEFGLQTTIDVLRKLEAAEAASEA-----LNNEKAANYKIEGLTEE 479
Qy 206 --KLIPEL-----KNKLKSVQNFHTLNTVKQANKDIDAAKLLTTE-----IAAI 250
Db 480 NVKLISELNETRDREKEKRAVEDLTAALS---EESDKAKEAHERYLSKEDDHEHALAQI 536
Qy 251 GEIKTETETTR---FVVDYDDMLSLLEAKAKMINTCNEYOKR-HGKKT 296
Db 537 GDLKALKSKTESYEVMLDEANYDITCLRNVDKLEAEVNVKRECESEKET 587

RESULT 7
US-10-080-608A-66
; Sequence 66, Application US/10080608A
; Publication No. US2003019956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-080-608A-66

Query Match      7.8%; Score 118; DB 14; Length 961;
Best Local Similarity 22.9%; Pred. No. 1;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;

Qy 7 DKTVEVVKNAIETADGALDY-NKYLDPVPMQTFDTIKELSRFKQYSAASVLVGD 65
Db 635 DKKEEVKKTLEQHDHSIVTHYKNMIREQDLQLEELKQOISTLKQNEQLQTAVTQVSO 694
Qy 66 K-----TLMDSDQK--YFEATQTVYEWGCVATQLLAAVILLFDEYNEKAS-- 110
Db 695 QQHKDQVNLKVLQKDSQHQGPYTDGAQ-----MNGVQPEISR--LREEIEELKSNRE 747
Qy 111 -----AQKDILIKVLDDGITK--LNEAQKSLVSSQFNASGKLALDLSQTLNDFSEK 162
Db 748 LLOSQAELKDSLLENKSSQLSPGTNEQSATAGDSEQIAELKQELATLKSQ--NSQSV 806
Qy 163 SSYFQSQVDKIRKEAYAGAAAGVAGPFGILISYSIAAGVVEGKL-----IPELKNK 216
Db 807 ITKLQTEKQELLQKTEAFKAPVGESETVIATKTTD--VEGRLSALLQETKELKNEIK 864
Qy 217 SVQNFHTLNTVKQANKDIDAAKLLTTEIAAIGIKTETETTRFYVDYDMLSLLE 276
Db 865 ALSEERTAIKEQLDSSNS-----TIALQNEKNKLEVDITDSKKEQDOLLV-LLAD 914
Qy 277 AAKMINTCNEYOK 290
Db 915 QDQKIFSLKNKLE 928

US-10-767-701-43815
; Sequence 43815, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 43815
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C3093_1.pcp
US-10-767-701-43815

Query Match      7.8%; Score 117.5; DB 16; Length 496;
Best Local Similarity 20.2%; Pred. No. 0.45;
Matches 68; Conservative 49; Mismatches 134; Indels 85; Gaps 10;

Qy 32 DQVPMQTFDE-----TKELSRFKQYSAASVLVGDITKLMDSDQKYEATQTVYEW 87
Db 76 DKATDHQOIEESLSRSTISEL-KVVQEALELSKSOVADLEQKLSAQDADISKLTEELNHC 134
Qy 88 GVATOLLAAVILLFDEYNEKASAKOILIKVLDDGITKLEAKQKSLVSSQFNASGK 147
Db 135 SSESLEKTKLLETETLTALLEELOAKLL--SLQEMETKLDQSKGRTSEALEKONGQ 192
Qy 148 LLALDLSQTLN-----DFSEKSSYFQSQVDKIRK----- 175

```

```
Db 193 LIVLQABLDNLKDNFTLQSLADLNKFSKSMHLQADELAKAQLVISEALSQKEEL 252
Qy 176 -----EAYAGAAGVWAGPFGLLIISYSI-----AAGVVEGKLIPELK----- 212
Db 253 ELNLSLSEQHGESKA-----FGENASQKILELEAQVHAMHAAEALNLELKEAASVK 306
Qy 213 ---NKLKSVQNFFTLNTVQKANDIDAAKLKUTTEIAAIGEIKTETETTRFYVD-YDD 268
Db 307 AAENKSDLEQQLSEIENKLVASSEEIELKRIQOEAAVSAERGMQLEETMTSVEGYKE 366
Qy 269 LML-----SILKEAAKMTNCTNEYOKR 291
Db 367 KITELQSLSSSVSKNQLSQEVKELDKCSEHQEQ 402

RESULT 10
US-10-282-122A-70920
; Sequence 70920, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70920
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70920

Query Match 7.6%; Score 114.5; DB 15; Length 1189;
Best Local Similarity 20.1%; Pred. No. 2.6;
Matches 67; Conservative 45; Mismatches 96; Indels 125; Gaps 13;

Qy 5 VADTKVEVVK-----NAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFKOEYSQAA 58
Db 161 IIEESAGLVKTKKKAESIQKLDHEDNLNVRVEDILYDLEGRVPLKEEAIAKEYKQLS 220
```

```
Qy 59 S-----VLVGDIKTLMDSDQKYFEATQTVYEW-----GVATQLLAA 96
Db 221 KEMEQSIVIVTSDI-----DHYTEDNQLRDLNHLKSOQAQKEGQQAQINQLLQR 272
Qy 97 YILLFDEYNEKKASAKQDILIKVLDGIGITKLENAQKILLVSSQSFNNASGKLLALDSQLT 156
Db 273 Y-----KGRQON-----DYDIEKLN-----YELVKATENYQLSGKLNVLZEKK 313
Qy 157 ND-----FSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVGEKLIPE 210
Db 314 NQSETNARYEEELDNLESQIDSINKV-----AQNEKLLAD 349
Qy 211 LKN-----KLSKSVNFTTSLNTVQKANKDI----- 236
Db 350 LKNKQKOLNEKVOELESLLVSDQHDQKLEIKNSYITLMSQSDVNNDIRPLEHTINE 409
Qy 237 -DAKCLKLWTE-IAAIGEIK-----TETETTRFY 263
Db 410 NEAKSKRLDSRLVEAFNQLKDIQONITQTNKAY 442

RESULT 11
US-10-467-421-37
; Sequence 37, Application US/10467421
; Publication No. US20040116665A1
; GENERAL INFORMATION:
; APPLICANT: Berthet, Francois-Xavier Jacques
; APPLICANT: Denoel, Philippe
; APPLICANT: Neyt, Cecile Anne
; APPLICANT: Poolman, Jan
; APPLICANT: Thonhard, Joelle
; TITLE OF INVENTION: Vaccine Composition
; FILE REFERENCE: B45259
; CURRENT APPLICATION NUMBER: US/10/467,421
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: PCT/EP02/01361
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: GB 0103171.5
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-467-421-37
```

```
Query Match 7.5%; Score 113; DB 16; Length 458;
Best Local Similarity 20.4%; Pred. No. 0.95;
Matches 68; Conservative 57; Mismatches 131; Indels 78; Gaps 12;

Qy 6 ADKTVVVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFKOE-----YSQA 57
Db 102 ATEKLEVAKEATQ-----DKVEKTSGLVEDIKDKAQSLOEADAADTVLEAKQA 148
Qy 58 ASVLVGDIKTLMDSDQKYFEATQTVYEW-----GVATQLLAAYILLFDEYN--- 105
Db 149 AS---DKVETTKAAEQSLKDDATQT-FESAKQAVEGKVEAIKEQVLQVDVSLKDDTDQDN 204
Qy 106 ----EKKASAKQDILIKVLDGIGITKLENAQKILLVSSQSFNNASG-----KLLALD--SQ 154
Db 205 TDQDQEKQTLKDAVQAATAAKRKVEDVDVVKHTTESFKNTASEKIDEIKQAADVKTTEE 264
Qy 155 LTNDSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVGEKLIPELKNK 214
Db 265 VKSLSQKADALKSSGSELKQTAQT-ANDAITFAQAQAVVSGVAAADSQAQSAKDK 323
Qy 215 LKSVQNFFTTSLNTVQKANKDIDAAKLKLTETIAAIGEIKTETETTRFYVDYDMLSL 274
Db 324 L-----NQLFEQKGSALDKVQELGEYGA-TEKINAVSEYVDLATQVI 365
Qy 275 KEAKGMINTC-----NEYQKRGHKTL 297
```

Db 366 KEEAQLQTNQAESLQAQAAGEYDATHEDKGL 399

RESULT 12

US-10-282-122A-58016
; Sequence 58016, Application US/10282122A
; Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patentin version 3.1

SEQ ID NO 58016

LENGTH: 1020

TYPE: PRT

ORGANISM: Enterococcus faecium

US-10-282-122A-58016

Query Match 7.5%; Score 112.5; DB 15; Length 1020;

Best Local Similarity 20.5%; Pred. No. 3.1;

Matches 64; Conservative 46; Mismatches 105; Indels 97; Gaps 11;

18 ETADG---ALDLNKNYLDQVPMQTFDETIKLSRQKQYSA-ASVLVGDIKTLIMDSQ 73

183 ETDENLSRVODITHELEEQTFPLAAQSEAKFELRLKETLTQTDVSLWVAEIKTKAD-- 240

74 DKYFEATQTVYECGVATQLAAAYILLFDEYNEKKASAKDILIKVLDDGKITLNFQKS 133

241 -----NDNKQAQL-----AKFNLELGLKLSI-----QEQES 267

134 LTVSSQFNNAASKLLALDSQTLNDRSEKSSYFQSQVDKIRKEAYAGAAAGVAGPGLI 193

268 ILAKQRKENAQADRLTEKNOQVLLDSEKLUKQTEGQKQVLDQET----- 311

194 ISVSIAAGVVEGKLIPELKNKLSVQNFPTTLNTVTKQA-----NKDIDAQKLKLTTEI 247

312 -----KHTQKSSQEQYQSLAEAQKKVKHFEKQLQESLMKAAAEKETEI 353

QY 248 --AAIGETKETE-----TTFYVDYDDMLMLSLKEPAKMMINTCNEYOKRH----- 292

Db 354 QKAEANLIKTOELEKQYKSTKELLAELRDQYVDLMQEA--AVGNELKYLEROYIOETA 411

QY 293 -GKTLFEVPEV 303

Db 412 KSKQTLAKQSEV 423

RESULT 13

US-10-369-493-22080

; Sequence 22080, Application US/10369493

; Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 22080

LENGTH: 1679

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-10-369-493-22080

Query Match 7.5%; Score 112.5; DB 14; Length 1679;

Best Local Similarity 20.6%; Pred. No. 6;

Matches 51; Conservative 44; Mismatches 67; Indels 85; Gaps 10;

QY 39 TPDETIKELSR-----FKQYSAASVLVGDIKTLIMDSQDYFEATQTVYECGVATQLL 94

Db 1134 SLEKTIQTLSEKEY-QCSAVIIDFKDITK-----EVTQV----- 1173

QY 95 AAVILLFDEYNEKKASAKQDI-----LIKVLDDGKITLNEAKSLVSSQSFNNAS 145

Db 1174 -----NILKENNAIILQKSLKNVTEKNREIYQLNDRQBEISRLQDLITQKEQVSINS 1226

QY 146 GKLLALDSQTL-----NDFSEKSSYFQSQ-VDKIRKEAYAGAAAGVAGPGLIISYSI 198

Db 1227 NKILVVESEMEQCKQRYQDLSQOQKDAQKDIKLTNE----- 1264

QY 199 AAGVVEGKLIPELKNKLSVQNFPTTLNTV-----KQANKDIDAQKLK-----LTTEIAAIG 251

Db 1265 -----ISDLKGLUSSAENANADLENKFNRLKQAEKLDASKKQQAALTNELNELK 1315

QY 252 EIKTETE 258

Db 1316 AIKDKLE 1322

RESULT 14

US-10-369-493-22285

; Sequence 22285, Application US/10369493

; Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

```
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22285
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22285

Query Match      7.4%; Score 112; DB 14; Length 1875;
Best Local Similarity 21.3%; Pred. No. 7.6;
Matches 68; Conservative 66; Mismatches 122; Indels 90; Gaps 15;

Qy 2 TEIVADTKVTVVVKNAITAGALDLYNK---YLDQVI-----PWQTFDETIK----- 45
Db 685 SRIIAERFKLLSNTLDTAENDQLRKFYDLQNTILKQDSKTHETLNEYVSCSKLSI 744
Qy 46 ---ELSRFKQY-----SQAAVLYGDIKTLMDSQDKYFEATQT 82
Db 745 VETELLNLKEEQKLRVHLEKNLQELNKLSPKDSLRIMVTQLQTKEREDLLEETRKS 804
Qy 83 VYEWCGVATQLLAAYILLFDEYN-----KKASAQKDLIKVL-DDGITTKLINEAQKSLV 136
Db 805 CQK-----KIDELDALSELKETSQKHIIKQLEEDNNSNIEWQNKIEA 850
Qy 137 SSQSFNNASGKLALDSQLTN-----DFSEKSSYFQSQVDKIRKEAYAGAAAGVVGPFGL 192
Db 851 LKQDYEV---ITSVDSKQTDIEKLQYKVSLEKEIEEDKIRLHTY-----NVMD---ET 899
Qy 193 IISVSIAGVVEGKL-IPELKNKLKSVQNFPTLNTVTKQANKDIDAAKLKLTTTIAAIG 251
Db 900 INDDSLRLEKSKINLTDAYSQIKYKDYETTSQSLQNTNSKLDSPKDFTNQIKNLT 959
Qy 252 EIKTETTRFYVDYDMLSLLEAAKMKMTNCNE--YORHGKK 295
Db 960 DEKTSLED-----KISLKE---QMFNLNLELDLQKGNEX 992

RESULT 15
US-10-369-493-5220
; Sequence 5220, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5220
; LENGTH: 2823
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5220

Query Match      7.4%; Score 111.5; DB 14; Length 2823;
Best Local Similarity 21.3%; Pred. No. 15;
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

Qy 25 DLYNKYLDQVTPWOT-FDETIKLSRFKQSVQAAASVLVG-----DIKTLMD 72
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADAERMSLVGKRINRYKEVSEIEKLARVEA 1939
Qy 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104
Db 1940 EDQIAYSRNSIEKARSEELMNFEDKEKINMTLAEPLDVEQCONITLL---YSQID 1996
Qy 105 NE-----KKASAQKDLIKVLDDGITTKLINEAQKSLVSS-----QSFNNAS--- 145
Db 1997 DEEYVQTAGRHAEKLEVQAO---KIVDRFVDTTETENPLKASHAYENIVEALKNATEAV 2053
Qy 146 -----GKLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVVGPFGLIIS 195
Db 2054 DSAEAASEAVSKMLGSESGDANEE---LRSQLEKLNES-----S 2094
Qy 196 YSIAAGVVEGKLIPELKNKLKSVQNFPTLNTVTKQANKDIDAAKLKLTTTIAAIGIKT 255
Db 2095 LSNVDNSNAVKIVEELKKEKKD-----LTDRLGHLN-----ELK-TSIVKRLGVKN 2140
Qy 256 ETETTRFYVDYDMLSLLEAAKMK-INTCNEYQKRGHKK 296
Db 2141 EASS---WDDKDRMHSILKNGAKTAHERSANVKESEGIKT 2179

RESULT 16
US-10-369-493-5221
; Sequence 5221, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5221
; LENGTH: 2823
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5221

Query Match      7.4%; Score 111.5; DB 14; Length 2823;
Best Local Similarity 21.3%; Pred. No. 15;
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

Qy 25 DLYNKYLDQVTPWOT-FDETIKLSRFKQSVQAAASVLVG-----DIKTLMD 72
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADAERMSLVGKRINRYKEVSEIEKLARVEA 1939
Qy 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104
Db 1940 EDQIAYSRNSIEKARSEELMNFEDKEKINMTLAEPLDVEQCONITLL---YSQID 1996
Qy 105 NE-----KKASAQKDLIKVLDDGITTKLINEAQKSLVSS-----QSFNNAS--- 145
Db 1997 DEEYVQTAGRHAEKLEVQAO---KIVDRFVDTTETENPLKASHAYENIVEALKNATEAV 2053
Qy 146 -----GKLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVVGPFGLIIS 195
Db 2054 DSAEAASEAVSKMLGSESGDANEE---LRSQLEKLNES-----S 2094
Qy 196 YSIAAGVVEGKLIPELKNKLKSVQNFPTLNTVTKQANKDIDAAKLKLTTTIAAIGIKT 255
Db 2095 LSNVDNSNAVKIVEELKKEKKD-----LTDRLGHLN-----ELK-TSIVKRLGVKN 2140
Qy 256 ETETTRFYVDYDMLSLLEAAKMK-INTCNEYQKRGHKK 296
Db 2141 EASS---WDDKDRMHSILKNGAKTAHERSANVKESEGIKT 2179

RESULT 17
US-10-369-493-5220
; Sequence 5220, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5220
; LENGTH: 2823
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5220

Query Match      7.4%; Score 111.5; DB 14; Length 2823;
Best Local Similarity 21.3%; Pred. No. 15;
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

Qy 25 DLYNKYLDQVTPWOT-FDETIKLSRFKQSVQAAASVLVG-----DIKTLMD 72
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKDADAERMSLVGKRINRYKEVSEIEKLARVEA 1939
Qy 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104
Db 1940 EDQIAYSRNSIEKARSEELMNFEDKEKINMTLAEPLDVEQCONITLL---YSQID 1996
```

```

US-09-978-309A-80
; Sequence 80, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-978-309A-80

Query Match 7.4%; Score 111; DB 10; Length 435;
Best Local Similarity 20.5%; Pred. No. 1.3; 108; Indels 62; Gaps 15;
Matches 60; Conservative 63; Mismatches 63;

QY 1 MTEIVADKTVEVWVNAIETADGALDLVNYKLDQVIMQTFDETTELKELSRFKQEYSQAASV 60
DB 99 LDNLLREKEVELEKHIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVOEKYNDTAQS 155
QY 61 LVGDIKTLLMDSQDKYFEATQTVVWCGVATOLLAAYILLFDEYNKKASAKDKILIKVL 120
DB 156 L-RDVTQALESQEKEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQSLRDVTAQ-L 205
QY 121 DDGITKLEAKSKLL-VSSQFNNAAGKLLAL-DSQLTN-DPSEKSSYFQSQVDKIRKEA 177
DB 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLLENLTLOEKVAMAESVEDVQQOI 265
QY 178 YAGAAAGVAGPFGILIIISYTAAGVVEGKLIPELKNK-----LKSQNEF---TTLS 226
DB 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEKEITETSSFLEKITDLK 308
QY 227 NTVKQANKDI-----DAKLKLTAIAAIGEIKTETETTFYVYDIDL 269
DB 309 NOLRQODEDFRKQLEKKGKRTAEKNVMTLT-----MEINKWRLLYEEL 353

RESULT 18
US-09-978-309A-77
; Sequence 77, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84

```

Db 583 LTBSTNQ-----EYA-----RMVQDLQNRSTLKEBEIKETSFLEKITDLK 625

Qy 227 NTVKQANKDI-----DAKLUKLTTEIAAIGETETETTRFRVYDDLL 269

Db 626 NQLRQODEDFRQLEKRGKRTABKNVMTLT-----MEINKWRLLYEEL 670

RESULT 20

US-10-282-122A-71606

; Sequence 71606, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 71606

; LENGTH: 938

; TYPE: PRT

; ORGANISM: Staphylococcus haemolyticus

US-10-282-122A-71606

Query Match 7.4%; Score 111; DB 15; Length 938;

Best Local Similarity 18.6%; Pred. No. 3.6;

Matches 60; Conservative 64; Mismatches 110; Indels 88; Gaps 11;

Qy 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFQKQYQAASVLVGDITKL 68

Db 189 TINKKNVAVAAHDSLPQINKIADRI---EYLDHDDDLKDYANQFR-----ALGNYKGD 240

Qy 69 LMDSDQKVFQATQVYEWCGVATQLLAAYILLDFEY-----NEKKAQAK--- 113

Db 241 ILDRQQKLNVDNAAI-----PSLNEKAKLILALNEYMPNIEKLDVANSIDIPQFPKINR 295

Qy 114 --DILIKVDDGIIKLENAQKSLVSSQ---SFFNASGKLLALDSQITNDFSEKSS--YF 166

Db 296 GVDIASEGFDLNRINDAQQYLTSAQORVGDYQEAAGRAQEVNQNANSALRQOSTGLP 355

Qy 167 OSQVDKIRKAYAGAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNKLSQVNFPTTLS 226

Db 356 QYQIKLSTD-----NSQDTVNDNQIVSN 379

Qy 227 NTVKQANKDIDAAKLMT-----TEIAAIGETETETTRFRVYDDLLMSL-- 273

Db 380 NDVKSMNSALAEALLTLSSNSDQAKATQSDIKALADISYGVIGSNRPTSFENDMLRNKLT 439

Qy 274 -LKEAAK---KMINTCNEYQKR 291

Db 440 RLENSSKSNQQLIDVLKLEKR 461

RESULT 21

US-09-884-696-5

; Sequence 5, Application US/09884696

; Publication No. US20030035809A1

; GENERAL INFORMATION:

; APPLICANT: GEORGE, LISLE W

; APPLICANT: ANGELOS, JOHN A

; APPLICANT: HESS, JOHN F

; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES

; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA

; TITLE OF INVENTION: BOVIS INFECTIONS

; FILE REFERENCE: 481.06

; CURRENT APPLICATION NUMBER: US/09/884,696

; CURRENT FILING DATE: 2001-06-19

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-884-696-5

Query Match 7.4%; Score 111; DB 10; Length 1023;

Best Local Similarity 21.3%; Pred. No. 4.1;

Matches 71; Conservative 56; Mismatches 116; Indels 90; Gaps 11;

Qy 2 TEIVADKTVEVVK-----NAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFQK 53

Db 155 TALSSMKIDELIKKQKSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207

Qy 54 YSOAASVL-----VGDIKTLMDSDQKVFQATQVYEWCGVATQLLAAYILL-FDEY 104

Db 208 LNKGLSVLSNTHLNGVGN-KLQNLPLNDINIGAGLDTV---SGILSAISAFILSNADAD 263

Qy 105 NEKKAQAKDILIKVLDD---GITKLENAQKSLVSSQSFNNASGKLLALDSQITNDFSE 161

Db 264 TGTAAAGVELTITKVLGNVGKISQYIIAQAAGLST----- 302

Qy 162 KSSYFQSQVDKIRKAYAGAAAGVAGPFGLLIIS---YSIAAGVVEGKLIPELKNKLS 217

Db 303 -----AAAAGLIASVVTLAISPLSFLSIADKFRANKIEEVSQRFKK 344

Qy 218 VQNFPTTLSNTVKQANKDIDAAKLMTTEIAAIGETETETTRF-----YVDYDDL 270

Db 345 LGYDGDLSLAAFHKETGAIDASLTRISTVLASVSSGISAATTSVGAPVSVLVGAVTGI 404

Qy 271 LSLKEAAK-----KMINTCNEYQKRHK 294

Db 405 ISGILEASKQMPFHVASKMADVIAEWKKGK 437

RESULT 22

US-10-032-585-7519

; Sequence 7519, Application US/10032585

; Publication No. US2003018093A1

; GENERAL INFORMATION:

; APPLICANT: Terry, Rosmer D.

; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone

; APPLICANT: Howard, Bussey

;; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;; FILE REFERENCE: 10182-005-999
;; CURRENT APPLICATION NUMBER: US/10/032,585
;; CURRENT FILING DATE: 2001-12-20
;; NUMBER OF SEQ ID NOS: 8000
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 7519
;; LENGTH: 1171
;; TYPE: PRT
;; ORGANISM: Candida albicans
US-10-032-585-7519

Query Match 7.4%; Score 111; DB 14; Length 1171;
Best Local Similarity 18.3%; Pred. No. 4.9;
Matches 40; Conservative 53; Mismatches 87; Indels 38; Gaps 5;
QY 103 EYNEKASAKQDILIK-VLDGGITKLNBAQKSLVSSQFNASGKLLALDSQLTNDFS 160
DB 289 EENLKTGSMKELEIKENQLNSDLRLNTARDIANDLNTTEKNKHTKLIQLSQIKQOLA 348
QY 161 EKSSVFQSQVDKIR-----KEAVAGAAAGVVGPFGLIISYSIAAGVVEGKLIPEL 211
DB 349 SNOQTFDQNEKYKQSNQELVQLKEEYANKQELLSTLSTGISTGNVTSGVI--TQLNDV 406
QY 212 KNLKSVQNFHTLNTVQKANKOIDAAKLTTEIAAIGIKTETETTRFYVD-----265
DB 407 KSKLASNFIKTSTLTKINHLOQOQMESQDKLT-----KAKTENETLLSAIBLHRQFI 459
QY 266 -----YDMLSLKKEAAKMMINTCNEYOKR 291
DB 460 KTKQLEIDSCLGFEPTKIHQLRDQSELISQNKFNQ 497

RESULT 23
US-10-282-122A-51864
;; Sequence 51864, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangsu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 51864
;; LENGTH: 1163
;; TYPE: PRT
;; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51864

Query Match 7.3%; Score 110; DB 15; Length 1163;
Best Local Similarity 17.9%; Pred. No. 5.9; Indels 152; Gaps 14;
Matches 73; Conservative 69; Mismatches 114;
QY 1 MTEIVADKTVEVKNATETADG-----ALDLYNKYLDQVIPWOTFDETIK-----45
DB 109 IVDITGDB-VEVLEGAKSVNEKQEIIGLSLDDFTRTV--VLPQGFSEFLKLEGKERR 165
QY 46 -----ELSRFKQEYSQAASVLVGDIK-----TLMDSQ 73
DB 166 NMLERLNLQYGDLSFLKARKIRKEREKENVLVGLKGYENINEDVLKERRELLKENN 225
QY 74 DKYFEATQTVVWCGVATQLAAAYILLFDEYN-----EKKASAKQDILIKV--119
DB 226 DFFNEASK-----EYLKAESEYNEGKEVWGLQIEIEEKURVRKDLMEKKDE 271
QY 120 --LDGGITKLNBAQKSLVSSQFNASGKLLALDSQLTN-----DSEKSS 164
DB 272 IDLKEKRALGESSKVKPYIDNYENTLKDILKEQLSRENTWTKALSKEDEMEKKLS 331
QY 165 YFQSQVDKIRKEAYAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELKNKLSVQNFHT 224
DB 332 IARDNKEK-----ALPKFMKHHIILDAIKEXDLDLONIKLEKKRLQKIEK 377
QY 225 LS-----NTVQANKOIDAAKLTTEIAAIGIKTETE-----258
DB 378 LSLEASNKEELIKQNIKIDSLLTKIQNLESKIDNLKVPPEYKKNKINEGIFLLRNYDEKL 437
QY 259 -----TTRFYVDYD-----DLMLSLKKEAAKMMINTCNEYOKR 291
DB 438 KHKNLGLDCDFQVDFEAKSKKEMLFNKLSEERSKL-----DITYTKK 481

RESULT 24
US-10-425-114-60225
;; Sequence 60225, Application US/10425114
;; Publication No. US20040034888A1
;; GENERAL INFORMATION:
;; APPLICANT: Liu, Jingdong
;; APPLICANT: Zhou, Yihua
;; APPLICANT: Kovalic, David K.
;; APPLICANT: Screen, Steven E.
;; APPLICANT: Tabaska, Jack E.
;; APPLICANT: Cao, Yongwei
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;; FILE REFERENCE: 38-21(53313)B
;; CURRENT APPLICATION NUMBER: US/10/425,114
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 73128
;; SEQ ID NO 60225
;; LENGTH: 420
;; TYPE: PRT
;; ORGANISM: Zea mays
;; FEATURE:
;; OTHER INFORMATION: Clone ID: LIB3607-037-H5_FLI.pep
US-10-425-114-60225

Query Match 7.3%; Score 109.5; DB 15; Length 420;
Best Local Similarity 19.1%; Pred. No. 1.6;
Matches 60; Conservative 82; Mismatches 103; Indels 69; Gaps 16;
QY 5 VADKTVEVKNATETAD--GALDLYNKYLDQVIPWOTFDETIKELSRFKQEYSQAASVL 61

Db 14 ISDRIAELDELEQARQQIALEEVKNHCO-----CEKLESLKQVNEKEDLVAAILAS 69
Qy 62 VGDIKTLMD--SODKYFEATQTVY-----EWCGVATQLLAAVILLPDEYNEKKASAKD 114
Db 70 KNEVEDLKGDMVSAAKHFEA-QLVHRDHEIEKCKQEAQVS-----EKYFHEKSTLESE 122
Qy 115 I-----LIKVLDDGIIKLN-----BAQ-KSLIVSSQSFNNASGKLLA-----LDSQL 155
Db 123 IORIQVVKVFEENLTQVADEKQLEAQLEQTSNDLDDSSAEIIKQLEIITKOLQERL 162
Qy 156 TNPSEKSSVFO--SOVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEGKLIPLKN 213
Db 183 ENDNEKSVLEERAMELEQVRQLEDSEASME-----LQTTIKNLQVLOEK--AELEN 236
Qy 214 KLSVQVNFVFTLNTVQKNDIAAKLUTTEIAAIGEIKTETETTRFFVYDDMLSL 273
Db 237 RMKDAE-----QATSDLSNLVASLDGKLK-----TEAQLQOLHAERAEATLE- 279
Qy 274 LKZAAKKNWNTCNE 287
Db 280 ----SEKQVSELNQ 289

RESULT 25
US-09-841-260-139
; Sequence 139, Application US/09841260
; Publication No. US20030175700A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; APPLICANT: Stromberg, Erika Jean
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS
; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515
; CURRENT APPLICATION NUMBER: US/09/841.260
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 140
; SEQ ID NO 139
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-260-139

Query Match 7.3%; Score 109.5; DB 10; Length 660;
Best Local Similarity 20.6%; Pred. No. 3;
Matches 65; Conservative 56; Mismatches 111; Indels 83; Gaps 15;

Qy 10 VEVVKNAIETADGALDLYNKYL-----DQV-----IPWQTFDE 42
Db 157 VNNIKALEAQKOTIDKLNKLVTLQNKSLTEVLKTTDSADQIPAINSOLEINKNSADQ 216
Qy 43 TIKELSRFKQEYSQAASVVLGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAVILLFD 102
Db 217 IIKDLERQNTSY-EAVLTNAGEVIKASSEAGIKLGQALQSIDV---AGDQSQAAVLQAAQ 272
Qy 103 EYNEKASAKQDILIKVLDGDIKTLNEAQKSLVSSQSFNNASGKLLALDSQLTNDPSEK 162
Db 273 NNSPDNIAATKEL-----IDAAETKVNLKQ-----EHTG---LTDSPLVKKAEEQ 315
Qy 163 SSVFQSQVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEGK-----LIPELKNK 214
Db 316 ISQAQKDIQEIKP---SGSDPIV-GPSSGAASAGSAGALKSSNNNSGRISLLDDVDNE 371
Qy 215 LKSV--QNFPTLNTVQKQAN-----KIDAAKLKLTTE---IAAIGEIKTETETTR 261
Db 372 MAALALQGF---RSMIEQFVNNPNATAKELQAMEAQLTAMSDQLVGADGELPAEIQAIK 427
Qy 262 FYVDYDDLMLSLKE 276
Db 428 -----DALAQALKQ 436

RESULT 26
US-10-007-693-139
; Sequence 139, Application US/10007693
; Publication No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; APPLICANT: Stromberg, Erika Jean
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND DIAGNOSIS
; TITLE OF INVENTION: OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007.693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 139
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-007-693-139

Query Match 7.3%; Score 109.5; DB 13; Length 660;
Best Local Similarity 20.6%; Pred. No. 3;
Matches 65; Conservative 56; Mismatches 111; Indels 83; Gaps 15;

Qy 10 VEVVKNAIETADGALDLYNKYL-----DQV-----IPWQTFDE 42

Wed Jan 5 14:16:57 2005

```

Db      157 VNNIKKALEAKQOTIDKLNKLVTLQNKQSLTEVLKTDSDAQIPAINSQLKNNKNSADQ 216
Qy      43 TIKELSRFQKQYQAAVILVGDITKLLMDSQDKYFEATQVYEWCGVATQLLAAVILLFD 102
Db      217 IIKDLERQNTISY-EAVLTNAGEVIKASSEAGIKLQALQSIVD---AGDSQAVLQAOQ 272
Qy      103 EYNEKASAKDILIKVLDGDKITKLENAQKSLIVSSQSFNNASGKLLALDSQLTNDFSEK 162
Db      273 NNSPDNIAATKEL-----IDAEYKVELKQ-----EHTG---LTDSPLVKKAEEQ 315
Qy      163 SSYFQSQVDKIRKAYAGAAVGVAGPGLIISVIAAGVVEGK-----LIPKLNK 214
Db      316 ISQAQKIOEIKP---SGSDIPIV-GPSSGAASAGSAGALKSSNNSGRISLLDLDVDNE 371
Qy      215 LKSV--QNFFTLSTNVKQAN-----KOIDAAKLKLTTE---IAAIGEIKTETETTR 261
Db      372 MAATIALQGF---RSMIEQFNNPNPATAKELQAMEAQUTAMSDQLVGADGELPAEIQAIK 427
Qy      262 FYVDYDDLLMLSLKE 276
Db      428 -----DALAQALKQ 436

RESULT 28
US-10-369-493-1061
; Sequence 1061, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1061
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-10-369-493-1061

Query Match      7.2%; Score 109; DB 14; Length 1005;
Best Local Similarity 19.1%; Pred. No. 5.8;
Matches 58; Conservative 55; Mismatches 93; Indels 98; Gaps 11;

Qy      7 DKTVEVVKNAIETADGALDLYNKLDQVTPWQTFDE---TIKELSRF-----KQYEQ 56
Db      605 DEILEDIKSQLNFKP---NPFYQYLSAVSYLNSVDEEGRNRIKEINISVGNKKEKRE 661
Qy      57 AASVL-----VGDITKLLMDSQDKYFEATQVYEWCGVATQLLAAVILLDFEYNEKKA 110
Db      662 ELNKLREDEIREINLKKLNELKKE-----LIEIENRRSLKFKYKYLGL 710
Qy      111 AQKDILIKVLDGDKITKLENAQKSLIVSSQSFNNASGKLLALD----- 152
Db      711 TEKLEELKNIKDGLIEI-----YNTCSKILAINIKKYNKEDIYVYLNK 757
Qy      153 -----SOLTNDPSEKSSYFQSQVDKIRKAYAGAAVGVAGPGLIISVIAAGVVEGKI 208
Db      758 ILEVNKEINDEERISYNQKLE-----INYN-----EEBHKKI 792
Qy      209 PEL-----KNKLSVQNFFTLSTNVKQANKDIDAAKLKLTTEAAGEIKTETETTRFYV 264
Db      793 KELYENKQELNDVREQKTEIETGYEYLLKQVESLKARL-----KEMENLEKEKEKLTQV 848
Qy      265 DYDD 268

```

```

Db      849 EYLD 852

RESULT 29
US-09-978-309A-79
; Sequence 79, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-978-309A-79

Query Match      7.2%; Score 108.5; DB 10; Length 476;
Best Local Similarity 20.2%; Pred. No. 2.3;
Matches 59; Conservative 63; Mismatches 109; Indels 61; Gaps 15;

Qy      1 MTEIVADKTEVVVKNATETADGALDLYNKLDQVTPWQTFDETIKELSRFQKQYQAAV 60
Db      99 LDNLREKEVELEKHAHAQAAILTAQEKYNTA---QSLRDVTAQLESVQEKYNTAQ 155
Qy      61 LVGDIKTLMDSQDKYFEATQVYEWCGVATQLLAAVILLDFEYNEKKAQKIDILIKVL 120
Db      156 L-RDVTAQLESQEKYNTAQSLRD---VTAQLESQ---EKYND-TAQSLRDVTAQ-L 205
Qy      121 DGGITKLENAQKSL--VSSQSFNNASGKLLAL--DSOLTN-DFSEKSSYFQSQVDKIRKEA 177
Db      206 ESVOEKYNTAQSLRDVTAQLESYKSTLKEIEDLKLNLTLQEKVMAEKSVEDVQOQI 265
Qy      178 YAGAAAGVAGPGLIISVIAAGVVEGKLIPELKNK-----LKSQNF----FTLSN 227
Db      266 LTAESTNQ-----EYA-----RMVQDLQNRSLKKEBEIKETSFLKIDTLKN 308
Qy      228 TVQKANKDI-----DAAKLKLTEAAGEIKTETETTRFYVDYDDL 269
Db      309 QLROQDEDFRKLBEKGRTAENNVNTELT-----MEINKWLLLYEEL 352

RESULT 30
US-10-739-930-6691
; Sequence 6691, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6691
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```



```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazov, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 10720
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11638C.1.pep
US-10-437-963-10720

Query Match          7.2%; Score 108; DB 16; Length 737;
Best Local Similarity 22.0%; Pred. No. 4.6;
Matches 71; Conservative 63; Mismatches 126; Indels 62; Gaps 14;

QY 4 IVADKTVEVVKNAIETAD---GALDLYNKYLDQVLPWQTFDETTELKLSRFKQEYSQ--- 56
DB 249 LVASQKLNICBAEIERLKWELGALTEANAAKAFDTQN--EITKELEDLTKLEEIKTN 307
QY 57 --AASVLVDGIKTLMSQDKYFQATQVYEWGCVATQLAAAYILLDFEYNEKKASAKQD 114
DB 308 KDLAESNGKLSRSELLSAEEKYSQSEAEV-----KYLQVMGA-VVEAKAAKAFAAEKE 362
QY 115 ILIKVLDGIIKLNAAOKSLVSSQFNNSGKLLALDSLTNDPSEKSSYFQSQVDKIR 174
DB 363 DIMKESDNLKRVKXIEIQSKLVSESDLELSEILSM-----KQKHGQFEVETSLK 414
QY 175 KEAVAGAAAGVVGPFGLIISYSIAAGVVEG---KLIPELKNLKSQVN----- 220
DB 415 KELGALAEKEI-----TTKAPVEVTEILKELEDLKRKVVEIQTNKDLVEVGND 464
QY 221 --FTTLNTVTKQANKOIDAQKLKLTETIAAIGIKETETETTRFVVDYDDLMLSL-LKEA 277
DB 465 KRLDLVSAQKQKQSIILEVEANNLKM--ELGALVEAK-EVATKAP--DAEKAKITKELEDV 519
QY 278 AKKMTNCNEKVRHGKKTLP 299
DB 520 KKKM-----EEIQKKDLVE 534

RESULT 34
US-10-205-647A-6
; Sequence 6, Application US/10205647A
; Publication No. US20040010812A1
; GENERAL INFORMATION:
; APPLICANT: University of Manitoba
; APPLICANT: Manitoba Cancer Treatment and Research Foundation
; APPLICANT: TURLEY, Eva A.
; APPLICANT: ENTWISTLE, Joycelyn
; TITLE OF INVENTION: HUMAN HYALURONAN RECEPTOR
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; STREET: 181 Freedman Crescent, Room 361
; CITY: Winnipeg
; STATE: Manitoba
; COUNTRY: Canada
; ZIP: R3T 5V4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/10/205,647A
; FILING DATE: 23-JULY-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CA97/00240
; FILING DATE: 10-APR-1996
; APPLICATION NUMBER: GB 9607441.4
; FILING DATE: 10
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-10-205-647A-6

Query Match          7.1%; Score 107.5; DB 15; Length 630;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 58; Conservative 65; Mismatches 107; Indels 63; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETTELKLSRFKQEYSQAASV 60
DB 253 LONLLREKEVELEKHTAAHAQAIIILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309
QY 61 LVGDITKTLMSQDKYFQATQVYEWGCVATQLAAAYILLDFEYNEKKASAKQDILIKVL 120
DB 310 L-RDVTAAQLESQEKYNDTAQSLRD---VTAQLESQ-----EKYND--TAQSLRDVTAQL- 359
QY 121 DGGITKLNAAQSKLL-VSSQSFNNSGKLLAL-DSOLTN-DFSEKSSYFQSQVDKIRKEA 177
DB 360 -ESQEKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLENTLQEKVMAEKSVEDVQOQI 418
QY 178 YAGAAAGVVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226
DB 419 LTAESTNQ-----EVA-----RMVODLQNRSLKEEIKETITSSFLEKITDLK 461
QY 227 NTVKQANKOI-----DAAKLTETIAAIGIKETETETTRFVVDYDDL 269
DB 462 NLQRQDEDFRKOLEKGRKTAENVMTELT-----MEINKWLLLYEEL 506

RESULT 35
US-09-978-309A-48
; Sequence 48, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; PRIOR FILING DATE: 1999-04-01
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-978-309A-48

Query Match          7.1%; Score 107.5; DB 10; Length 631;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 58; Conservative 65; Mismatches 107; Indels 63; Gaps 15;
```


RESULT 40
US-10-028-248A-105
; Sequence 105, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Sureesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 105
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-028-248A-105
Query Match 7.0%; Score 106; DB 14; Length 1961;
Best Local Similarity 20.4%; Pred. No. 25;
Matches 69; Conservative 69; Mismatches 120; Indels 80; Gaps 16;
Qy 3 EIVADKTVVVKNAIETADGALDLYNKYLDQV-IPWQTFDETIK-----ELSRFKQEVSOA 57
Db 1140 ELEALKT--ELEDTLDTAAQQLSRKQEVSVILKKTLEDEAKTHPAQIQEMRQKHSOA 1197
Qy 58 ASVLVGDKITLMDSDQKYFEAT-----QTVYECG-VATQLLAAYILL-----FDEYNEK 107
Db 1198 VEELAEQIE-----QTKRVKATLEKAKQTLNERGELANEVKA---LLQKGDSEHKRK 1248
Qy 108 KASAQ-KDILIKV-----LDDGITKLENAQKSLVSSQSFNNASGKIL-----ALDS 153
Db 1249 KVEAQLQELQVFKFGEGRVRLTELADKYSKLQVELDSVTGLLNQSDSKSKLTKDFSALES 1308
Qy 154 QLTN-----DFSEKSSYFQSQVDKIRKEAYAGAAAGVAGPGLI 193

Db 1309 QLODTQELLOENRQKLSLSTKLQMEDEKNSFRELEEEERAKRN----- 1355
Qy 194 ISYSIAAGVVEGKLIPELKNKJKSVQNFPTTSLNTVQKANKDIDAAKLKTETETAAIGBI 253
Db 1356 LEKQIATLHAQ---VTDMKKKMGDVGVCLETAEEAKERLQKDLGLSQRLEEKVAAYD- 1410
Qy 254 KTETETRTFYVDYDDMLSLKKEAAKMMINTCNEYQKR 291
Db 1411 KLEKTKTRLQOELDDLVL--DHQRQSVSNLEKKQKK 1446
RESULT 41
US-10-107-782-105
; Sequence 105, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gangolli, Esha
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Malyankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Millet, Isabelle
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Sciore, Paul
; APPLICANT: Shenoy, Sureesh
; APPLICANT: Shimkets, Richard
; APPLICANT: Si, Jingsheng
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly
; APPLICANT: Stone, David
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 105
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-105
Query Match 7.0%; Score 106; DB 15; Length 1961;
Best Local Similarity 20.4%; Pred. No. 25;
Matches 69; Conservative 69; Mismatches 120; Indels 80; Gaps 16;

[illegible]

```

RESULT 42
US-09-815-242-5815
; Sequence 5815, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5815
; LENGTH: 2076
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5815

```

```

QY 66 KTLIMDSQKYPFATQTVYVWCVAQTLAAAYILLFDEYNEKKASAKOILIKVLDDGIT 125
Db 1727 KVLIPQVQVKP-AAROSV-----GVKAEAQNALIDQSDLSSTEEBELAAKHLVEQALNQAID 1781
QY 126 KLNBAOKSLIVSSQSFN--NASGK-----LLALDSQITNDFS-----EK 162
Db 1782 QIHADKTAQVNODSINAQNIISKIRPATVTKATALQIQIQAATNLIKLIKANNEATDEE 1841
QY 163 SSYFQSOVDKIRKEAVAGAAAGVAGFPGLIISVSIAAGVVVEGKLIPELQNKLSKVQNF 222
Db 1842 QNTAIAQVEKELIKAKQOIASAVTNAD-----VAYLHDEKNEIRETEPVINRKASAREQL 1897
QY 223 TTLSNTVQKANKDIDRAKLLKTTT-----IAAIGEIKTB-----TETTR 261
Db 1998 TTLENDKKQALE---ANIQTAVEERNSILAQLNIYDTAIGQIDODRSNAQVMDKTASLN 1953
QY 262 FYVDYDMLSLKKEAAKMIN 283
Db 1954 LQTIHDLVHPPIKPPDAEKTIN 1975

RESULT 43
US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5639

```



```

RESULT 47
US-10-282-122A-53523
; Sequence 53523, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangseu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A

```

;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53221)B

;; CURRENT APPLICATION NUMBER: US/10/437,963

;; CURRENT FILING DATE: 2003-05-14

;; NUMBER OF SEQ ID NOS: 204966

;; SEQ ID NO 191043

;; LENGTH: 815

;; TYPE: PRT

;; ORGANISM: Oryza sativa

;; FEATURE:

;; OTHER INFORMATION: Clone ID: PAT_MRT4530_8739C.1.pap

US-10-437-963-191043

Query Match 7.08; Score 105.5; DB 16; Length 815;

Best Local Similarity 20.18; Pred. No. 8.5;

Matches 75; Conservative 68; Mismatches 131; Indels 99; Gaps 14;

QY 8 KTVGVVQNAJETA-----DGDALDYNK-YLDQVVPWQTFDETIKELSRFKQVYQAAASVLV 62

DB 216 KALESVVSSWESKLAENEAISLMREYERKLLDQQAQK--KQMKFQQAASLLDQLS 273

QY 63 GDIKTLMDSDQKYFEATQTVYWCQVATQLLAAAYILLPDEYNKKAQAQDILIKVLDD 122

DB 274 STKKTVTSLEE--FREKTLAEELREIRRL-----ESSLAQAQDDKDVLEAKLKE 323

QY 123 GITKLNKAQSLVSSQSFNNASGKLJLALDSQLT----- 156

DB 324 KLGDNVILQEKVLSLQSDIDNGIRIELSLSLSSKEADYRNLCSFSDQTKESIELAEAK 383

QY 157 -----NDFSEKSYFQSDVKIRKEAVAGAA-----GVVAGPGLIISY 196

DB 384 IOLEEEVHRTRNDLSKI-----SSIDLNEELQALNSAKNEAEKSELTKDYTDLUKAS 439

QY 197 SIAAGVVEGKLIPELKNKLSQVNFFT-TLSNTVKQ-----ANKOIDAAKLITTEIAA 249

DB 440 SEARESNSLELLEKNWIKQDGLSDALSQSKDRENTAALNKELDATKAMLENEVAA 499

QY 250 IGEIK-----TETTRFRVYD-----DMLSLLEKAAKQKMTNCEYQK 290

DB 500 VKSLRESLQSTEEALTSRSEVSKLSVELDEANRMDQLVLI-----SKLQDEFNEMQE 554

QY 291 RHGKKTLPFVPEV 303

DB 555 GLTNK-LGEVESV 566

RESULT 49

US-10-369-493-11332

Sequence 11332, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 11332

LENGTH: 873

TYPE: PRT

ORGANISM: Methanosarcina mazei

US-10-369-493-11332

Query Match 7.08; Score 105.5; DB 14; Length 873;

Best Local Similarity 20.87; Pred. No. 9.3;

Matches 77; Conservative 41; Mismatches 117; Indels 135; Gaps 15;

QY 5 VADKTVEVQNAJETADGALDLYNKYLDQV-----IPWQTFDETI-----KEL----- 47

DB 112 LTDKSVLGLANKVNDERFAYDCYRRFISMFGDVVLGIDFDKFEESLIDBKKEKELKVESDTD 171

QY 48 -----SREKQVYQAAASVLVGDIKTLMDSDQKYFEATQTVYWCQVATQLLAAAY 97

DB 172 LDKALKDLAERFK-----GVIKL-----EKGFEPQ-----DPKVQLQWAI 208

QY 98 ILLPDEYNKKAQAQDILIKVLDDGITKLNKAQK-----LLVSSQSF-----NNASGKLLA 150

DB 209 DAVFDSNNPRAITYR-----KLNEIDDSNGTAVNVQTMVYGNRGNTSGTGVA 256

QY 151 LDSQLTNDSEKSSYFQSDVKIRKEAYAGAAAGVAGPGLIISISYIAAGVVECKLIPE 210

DB 257 F---TRNPSTGERKFFGEYLINAQGE-----DVAAGIRTPDFIDT 293

QY 211 LKNKLSQVNFFTLSNTVKQANKDID----- 247

DB 294 LGNKITPEAYNQLVDICRLEAHFKMDQDIEFTQEGKLYMLQTRTGKRTAAAVKLAIDM 353

QY 248 AAIQEIKTETETTRFRVYDYLML-----SLLEKAAK-----KMINTCN--E 287

DB 354 VAESGLDKETAVTRVKAEDHIDLLHPRIDPNKLEVVAKGLPASPGAAGVKVVFATSAEAE 413

QY 288 YQKRHGKKTLP 297

DB 414 EMAEKGKTI 423

RESULT 50

US-10-393-602-148

Sequence 148, Application US/10393602

Publication No. US20030170714A1

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

NUMBER OF SEQUENCES: 151

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/393,602

FILING DATE: 19-Mar-2003

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/592,126

FILING DATE: 26-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:

LENGTH: 1312 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

```

; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-393-602-148

Query Match      7.0%; Score 105.5; DB 14; Length 1312;
Best Local Similarity 18.8%; Pred. No. 16;
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;

Qy 10 VEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIK-----ELSRPKQEY 54
Db 502 MEVISLQNEKAD--LDTRLKLDQEMEQLNHHYTRTQMEMLTQKADKDEQIRKIKSRH 559
Qy 55 SQAASVLVGDIKTLTLLMDSQDKYFEATQTVYEWCGVATOLLAAYILLDFEYNEKKSAAQKD 114
Db 560 SDELTSLLG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 591
Qy 115 ILIKVLDDGITKLINEAOKSLIVSSQSFNNASGKLLALDLSOLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKNHNNELKRREQLSS--YEDKLFVCGSQDPFES 640
Qy 169 QVDKIRKE-----AYAGAAAGVVAGPF----- 190
Db 641 DLDRLKEEIEKSKORAMLAGATA--VYSQFITQITLDENQSCCPVCQVFTAEALQEV 698
Qy 191 -----GLI-ISYSIAAGVVEGKLIPELNKLS 217
Db 699 SDLQSKRLAPDKLKSTESLKKKRRDEMGLVPMRQSIID--LKEKEIPELNKLN 756
Qy 218 VQFFTLISNTVKQANKDIDAQKLKLTETAAGEIKTETTT-----RFYVDY 266
Db 757 V-----NRDIQRLKNDIBEQETLLGTIMPEESAKVCITDVTIMERFQNEL 802
Qy 267 DDLMLSLKKAAC-----KMINTCNE--YQKRHGKKT 297
Db 803 KQVERKIAQQAQKLGIDLDRTVQVQVQKQKQKQKLDTV 842
```

Search completed: January 5, 2005, 11:08:51
Job time : 66.2944 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:51:49 ; Search time 15.9474 Seconds
(without alignments)
1260.042 Million cell updates/sec

Title: US-09-993-292b-28

Perfect score: 1508

Sequence: 1 MTEIVADKTVVEVKNNAIETA.....TCNEYQKRGKTLFVPEV 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1499	99.4	309	1	Sequence 3, Appli
2	1499	99.4	309	5	Sequence 3, Appli
3	124	8.2	1036	4	Sequence 7736, Ap
4	118	7.8	961	4	Sequence 66, Appl
5	118	7.8	1211	3	Sequence 4820, Ap
6	116	7.7	885	4	Sequence 1660, Ap
7	116	7.7	1129	4	Sequence 8019, Ap
8	115.5	7.7	718	4	Sequence 2753, Ap
9	114	7.6	515	4	Sequence 5317, Ap
10	113.5	7.5	1196	4	Sequence 3944, Ap
11	112	7.4	606	4	Sequence 2, Appli
12	112	7.4	631	4	Sequence 11, Appl
13	111.5	7.4	808	4	Sequence 4995, Ap
14	111	7.4	477	1	Sequence 3, Appli
15	111	7.4	477	1	Sequence 3, Appli
16	111	7.4	477	3	Sequence 3, Appli
17	110.5	7.3	924	4	Sequence 18798, A
18	109.5	7.3	1454	4	Sequence 5793, Ap
19	108.5	7.2	565	4	Sequence 807, App
20	107.5	7.1	829	4	Sequence 20145, A
21	106.5	7.1	1231	4	Sequence 5150, Ap
22	105.5	7.0	1312	2	Sequence 148, App
23	105.5	7.0	1312	2	Sequence 51, Appl
24	105.5	7.0	1312	4	Sequence 148, App
25	104	6.9	794	4	Sequence 1050, Ap
26	104	6.9	962	3	Sequence 4497, Ap
27	103.5	6.9	349	4	Sequence 13939, A

Query Match 99.4%; Score 1499; DB 1; Length 309;

ALIGNMENTS

RESULT 1

US-08-557-115-3 Application US/08557115

Sequence 3, Appli

Patent No. 5731151

GENERAL INFORMATION:

APPLICANT: King, Harold C.

APPLICANT: Sathish, Mundayoor

APPLICANT: Shinnick, Thomas M.

TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysis

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.A.

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,115

FILING DATE: 26-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greene, Jamie L.

REGISTRATION NUMBER: 32,457

REFERENCE/DOCKET NUMBER: 03063-0171US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 818-3700

TELEFAX: (404) 818-3799

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-557-115-3

Best Local Similarity 100.0%; Pred. No. 2.3e-134;
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSAASV 60
Db 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSAASV 60
Qy 61 LVGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKIDILIKVL 120
Db 61 LVGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKIDILIKVL 120
Qy 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDSQTLNDFSEKSYFQSQVDKIRKEAYAG 180
Db 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDSQTLNDFSEKSYFQSQVDKIRKEAYAG 180
Qy 181 AAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTLNTVKQANKDIDAAK 240
Db 181 AAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTLNTVKQANKDIDAAK 240
Qy 241 LKLTETIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKQKMTNCNEYQKRHGKKTILFEV 300
Db 241 LKLTETIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKQKMTNCNEYQKRHGKKTILFEV 300
Qy 301 P 301
Db 301 P 301

RESULT 2
PCT-US94-05869-3
; Sequence 3, Application PC/TUS9405869
; GENERAL INFORMATION:
; APPLICANT: King, C. H.
; APPLICANT: Athish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIS
; TITLE OF INVENTION: OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 1200, The Candler Building
; STREET: 127 Peachtree Street, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05869
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US94-05869-3
Query Match 99.4%; Score 1499; DB 5; Length 309;
Best Local Similarity 100.0%; Pred. No. 2.3e-134;

Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSAASV 60
Db 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSAASV 60
Qy 61 LVGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKIDILIKVL 120
Db 61 LVGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAQKIDILIKVL 120
Qy 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDSQTLNDFSEKSYFQSQVDKIRKEAYAG 180
Db 121 DDGITKLNKQAQSLVSSQSFNNASGKLLALDSQTLNDFSEKSYFQSQVDKIRKEAYAG 180
Qy 181 AAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTLNTVKQANKDIDAAK 240
Db 181 AAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTLNTVKQANKDIDAAK 240
Qy 241 LKLTETIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKQKMTNCNEYQKRHGKKTILFEV 300
Db 241 LKLTETIAAIGEIKTETETTRFYVDYDDLMLSLKKAQKQKMTNCNEYQKRHGKKTILFEV 300
Qy 301 P 301
Db 301 P 301

RESULT 3
US-09-543-681A-7736
; Sequence 7736, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7736
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7736

Query Match 8.2%; Score 124; DB 4; Length 1036;
Best Local Similarity 21.5%; Pred. No. 0.0088;
Matches 82; Conservative 56; Mismatches 127; Indels 116; Gaps 16;
Qy 4 IVADKTVVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSAASV-- 61
Db 178 IKKQKNGENVSS--ELAKASIDILINQLVD-----TASSLNNNISAFSQQNLKGLSVLSN 230
Qy 62 -----VGDITKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILL--FDEYNEKKAQAQK 114
Db 231 TKHLNGVGN--KLQNLPLNDKLTGLDVT---SGILSAISAFSFLNSADADGTGKAAAGVE 286
Qy 115 ILIKVLDD-----GITKLNKQAQSLVSSQSFNNASGKLLALDSQ--T 156
Db 287 LTKVLGNVGKANSQYILAQVAGLS--TSAASAGLIASAVTLAISPLSLAIADQFKRA 345
Qy 157 NDFSEKSSYFQ-----SQVDKIRKEA-----YAGAAGVVGAGPFG 192
Db 346 NKIEYSQRPKKFGYEGDSLAAAFRKETGAIDASLTINTALGTISAGISAASTASLIGA 405
Qy 193 IISYSIAA--GVVEGKLIPELKNKLSQVQNFPTLNTV----- 229
Db 406 PISALVGAITGISGILEASKQSMFEHVN---RWNTIIAEWEKTHGKNFPENGYDARHS 462
Qy 230 -----KQANKDIDAAKILTT-----ETAAIGEIKTETETTRFYVDYDDLML 270

Db 463 AFLEDNFKLLSQYNKEYSVSVLITQOHWDLIGELASVTNGAKTLGSKSYIDY----- 518
Qy 271 LSLKBAAKMINTCNBYQKR 291
Db 519 -----YEEGKRLKKPNEFQK 535

RESULT 4

US-09-914-259-66
; Sequence 66, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-914-259-66

Query Match 7.8%; Score 118; DB 4; Length 961;
Best Local Similarity 22.9%; Pred. No. 0.029;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;
Qy 7 DKTVEVUNAETADGALDLY-NKYLQVTPMOTFDTETIKELSRFKQYEQAAVLVGD 65
Db 635 DKKEEVKTKLEQHDSTVTHYKNMIREQDLQLELKQKQISTLKQNEQLQTAVTQQVSOI 694
Qy 66 K-----TLLMSQDK--YFEATQTVYEWCGVATQLLAAAYILLDFEYNEKKAS-- 110
Db 695 QQHKDQVNLKVKQKDSQHQSGPTDGAQ-----MNGVQPEISR---LRBEIBELKSNRE 747
Qy 111 -----AQKQILIKVLDDGITK--LNEAQSLLVSSQSFNNASGKLLALDLSQLTNDPSEK 162
Db 748 LLQSLAEKSLIENAKSSQLSPCTNQSSATAGDSQIAELKQELATLKSQ-LNSQSV 806
Qy 163 SSYPQSQVDKIRKBAAGAAVGVAGPGLIISYIAAGVVEGKL-----IPELKNKLK 216
Db 807 ITKLQTEKQELLQKTEAFKAPVGESETVIATKTTD--VEGRLSALLQETKELKNEIK 864
Qy 217 SVQNFPTLSNTVQKANKDIDAAKLTTETAAIGEIKTETETTRFFVYVDYDMLSLIKE 276
Db 865 ALSEBTAIRKQDSSNS-----TIALQNEKNKLEVDITDSKKQBDDLLV-LLAD 914
Qy 277 AAKQMINTCNBYQK 290
Db 915 QDQKIFSLKNKLKE 928

RESULT 5

US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820

; LENGTH: 1211
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

Query Match 7.8%; Score 118; DB 3; Length 1211;
Best Local Similarity 19.3%; Pred. No. 0.041;
Matches 69; Conservative 53; Mismatches 112; Indels 124; Gaps 13;

Qy 5 VADTKVEVVK-----NAISTADGALDLYNKYLDQVTPMOTFDTETIKELSRFKQYEQAA 58
Db 183 IIEESAGVLKXKKAESIQKLDHTEDNLNRVEDILYDLGRVPEPLKEAAIAKEYKQLS 242
Qy 59 S-----VLVGDITKLLMSQD-----KYFEATQTVYEWCGVATQLLAAAYILLDFEYN 105
Db 243 KEMEQSDVITVTSDDIHYTEDNQRLDERLNLHLKSQQAQKE--GQQAQINQ-----LLOK 296
Qy 106 EKASAKQDILIKVLDDGITKLEAQSLLVSSQSFNNASGKLLALDLSQLTND-----F 159
Db 297 GKRQNDYD-----IEKLN---YELVKATENYEQLSGKLVLEERKKNQSETNARY 344
Qy 160 SEKSSYFQSQVDKIRKEAYAGAAAGVGVAGPGLIISYIAAGVVEGKLIPELKN----- 213
Db 345 EEELDNLESQIDSIKNEK-----AQNEKLLADLNKQKQLN 380
Qy 214 -----KLKSVQNFPTLSNTVQKANKDI-----DRAKLKLT 244
Db 381 KEVQELLESLLYISDEQHQDEKLEIKNSYITLMESEQSVVNDIRFLEHTINENEAKSRLD 440
Qy 245 TETAAGEIKTETETTRFFVYDYLMLSLKEAAKQMINTCNEYQKRGHKKTLPEVPE 302
Db 441 SRLVE-----APNQLKDIOQNTQTQKEYQS---SKSKSEKVEQ 476

RESULT 6

US-09-710-279-1660
; Sequence 1660, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1660

Query Match 7.7%; Score 116; DB 4; Length 885;
Best Local Similarity 20.0%; Pred. No. 0.04;
Matches 63; Conservative 42; Mismatches 92; Indels 118; Gaps 12;

Qy 42 ETIKELSRFKQYEQAAAS-----VLVGDITKLLMSQD-----KYFEATQTVYEWCG 88
Db 2 EPLKEAAIAKEYKQLSKEMEQQSDVITVTSDDIHYTEDNQRLDERLNLHLKSQQAQKE--G 59
Qy 89 VATQLLAAYILLDFEYNEKKASAKQDILIKVLDDGITKLEAQSLLVSSQSFNNASGKL 148
Db 60 QQAQINQ-----LLOKYGKQNDYD-----IEKLN---YELVKATENYEQLSGK 103
Qy 149 LALDLSQLTND-----FSEKSSYFQSQVDKIRKEAYAGAAAGVGVAGPGLIISYIAAGV 202
Db 104 NVLEERKKNQSETNARYEEELDNLESQIDSIKNEK-----A 139

[illegible]

APPLICATION NUMBER: US/08/477,831C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: PIERRI, MARGARET A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: SIM-10
TELEPHONE: 212-596-9090
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 606 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc = "RHAMM I protein"
US-08-477-831C-2

Query Match 7.4%; Score 112; DB 4; Length 606;
Best Local Similarity 20.5%; Pred. No. 0.034;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSQAASV 60
DB 228 LDNLLREKEVELEKHIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 284
QY 61 LVGDIKTLMDSDQKYFEATQTVYWCWGVATOLLAAYILLDFEYNEKASAKDILIKVL 120
DB 285 L-RDVTQAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 334
QY 121 DDGITKLNBAOKSL-L-VSSQSFNNASGKLLAL-DSOLTN-DFSEKSSYFQSQVDKIRKEA 177
DB 335 ESVOEKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLNLTLQEKVAMAEKSVEDVQOOI 394
QY 178 YAGAAAGVAVGPGFLIISYSIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226
DB 395 LTAESTNQ-----EYA-----RMVDLQNRSTLKEEIKETITSSFLEKITDLK 437
QY 227 NTVKQANKDI-----DRAKLTETAAIGETETETETETETETETETETETETETETET 269
DB 438 NQLRQODEDFRKLEKGRKTAENVMTELT-----MEINKWRLLYDEL 482

RESULT 12
US-08-477-831C-11
Sequence 11, Application US/08477831C
Patent No. 6429291
GENERAL INFORMATION:
APPLICANT: TURLEY, EVA A.
APPLICANT: SHUEN, ZHANG
APPLICANT: ENTWISTLE, JOYCELYN
TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10020-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Rel. #1.0, ASCII
CURRENT APPLICATION DATA: US/08/477,831C
APPLICATION NUMBER: US/08/477,831C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: PIERRI, MARGARET A.

REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: SIM-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9090
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc = "RHAMM I-2a"
US-08-477-831C-11

Query Match 7.4%; Score 112; DB 4; Length 631;
Best Local Similarity 20.5%; Pred. No. 0.058;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

QY 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQYSQAASV 60
DB 253 LDNLLREKEVELEKHIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309
QY 61 LVGDIKTLMDSDQKYFEATQTVYWCWGVATOLLAAYILLDFEYNEKASAKDILIKVL 120
DB 310 L-RDVTQAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 359
QY 121 DDGITKLNBAOKSL-L-VSSQSFNNASGKLLAL-DSOLTN-DFSEKSSYFQSQVDKIRKEA 177
DB 360 ESVOEKYNDTAQSLRDVTAQLESYKSTLKEIEDLKLNLTLQEKVAMAEKSVEDVQOOI 419
QY 178 YAGAAAGVAVGPGFLIISYSIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226
DB 420 LTAESTNQ-----EYA-----RMVDLQNRSTLKEEIKETITSSFLEKITDLK 462
QY 227 NTVKQANKDI-----DRAKLTETAAIGETETETETETETETETETETETETETETET 269
DB 463 NQLRQODEDFRKLEKGRKTAENVMTELT-----MEINKWRLLYDEL 507

RESULT 13
US-09-543-681A-4995
Sequence 4995, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4995
LENGTH: 808
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-4995

Query Match 7.4%; Score 111.5; DB 4; Length 808;
Best Local Similarity 22.0%; Pred. No. 0.093;
Matches 66; Conservative 55; Mismatches 108; Indels 71; Gaps 14;

QY 4 IVADKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQE-YSOAASVLV 62
DB 257 IIESEAKKIVKKIQTSTIQLNNAL--LASV---RTGRELHKYKQKCAFCGNILP 311
QY 63 GDIKTLMDSDQKYFEATQTVYWCWGVATOLLAAYILLDFEYNEKASAKDILIKVLDD 122
DB 312 ED---LQKIDKHFNKES-----EELHE-----SLNILLSTIEK 342
QY 123 GITKLNBAOKSL-L-VSSQSFNNASGKLLALD---SOLTNDFSEKSSYFQSQVDKIRKEAYA 179

Db 343 ELKIPNLK--IKTIDFYNFANDLNTLNDLSKSIDYKSLUSLSLKEQEKRKSDIFT 400
Qy 180 GAAAGVAGPGLIISYIAAGVVEGKLIPELKNKLSVQNFVFTTSLNT--VKQANKDID 237
Db 401 -----PLEFHSSVSVEDSINELRYSYE---KIRSKSNDPTKSLNTRQVKARN--- 444
Qy 238 AAKLKLTTAAIGETETTRFYVDYDDMLSLKBE---AAKQWINTCNEYQKRGK 294
Db 445 --ELRLHEVYKFTDIK-----YVDERNTINKLKEKEDLEKTSINIKQNVDEKRRK 494

RESULT 14

US-08-402-217A-3
; Sequence 3, Application US/08402217A
; Patent No. 5587301
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,217A
; FILING DATE: 10-MAR-1995
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PP-0028US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-0195
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; LIBRARY: mouse
; CLONE: GI 53979
; US-08-402-217A-3

Query Match 7.4%; Score 111; DB 1; Length 477;
Best Local Similarity 20.5%; Pred. No. 0.047;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQEYSQAASV 60
Db 99 LDNLREKEVELEKHAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
Qy 61 LVGDIKTLTLLMDSQDKYFEATQTVYVCGVATQLLAAAYILLFDFYNEKASAKQDKILIKVL 120
Db 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 205
Qy 121 DDGITKLINEAKQSL-L-VSSQSFNNASGKLIAL-DSQLTN-DFSEKSYFQSOVDKIRKEA 177
Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIDLKLNLTLQEKVMAEKSEVEDVQOQI 265

Qy 178 YAGAAAGVAGPGLIISYIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEBEIKETISFLEKITDLK 308
Qy 227 NTVKQANKDI-----DAAKKLTTTAAIGETETTRFYVDYDDDL 269
Db 309 NQURQDEDFRKQLEBKGRKTAENVMTELT-----MEINKWRLLYEEL 353

RESULT 15

US-08-700-178-3
; Sequence 3, Application US/08700178
; Patent No. 5783669
; Patent No. 5783669 5700912
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,178
; FILING DATE: August 20, 1996
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0028-1 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; LIBRARY: mouse
; CLONE: GI 53979
; US-08-700-178-3

Query Match 7.4%; Score 111; DB 1; Length 477;
Best Local Similarity 20.5%; Pred. No. 0.047;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQEYSQAASV 60
Db 99 LDNLREKEVELEKHAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
Qy 61 LVGDIKTLTLLMDSQDKYFEATQTVYVCGVATQLLAAAYILLFDFYNEKASAKQDKILIKVL 120
Db 156 L-RDVTQALESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 205
Qy 121 DDGITKLINEAKQSL-L-VSSQSFNNASGKLIAL-DSQLTN-DFSEKSYFQSOVDKIRKEA 177

Db 206 ESQKYNDAQSLRDVSAQLESYKSTLKEIEDLKLENLTLOEKVAMAEKSVEDVQQOI 265
Qy 178 YAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIEKITSFLEKITDLK 308
Qy 227 NTVKQANKDI-----DAAKLKLTTETIAAIGETKTETTRFRFYVDYDDL 269
Db 309 NQLRQODEFRKQLEBKGRKTAENVTMTLT-----MEINKWRLLYEEL 353

RESULT 16

US-08-995-654-3
; Sequence 3, Application US/08995654
; Patent No. 6025138
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig
; APPLICANT: Seilhamer, Jeffrey
; TITLE OF INVENTION: HVALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UNBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,654
; FILING DATE: December 22, 1997

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/700,178
; FILING DATE: August 20, 1996

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/402,217

; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0028-2 DIV
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166

; TELEX:
; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids

; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-995-654-3

Query Match 7.4%; Score 111; DB 3; Length 477;

Best Local Similarity 20.5%; Pred. No. 0.047;

Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 MTEIVADKTVVEVKNKAEITADGALDLYNKYLDQVTPWQTFDTIKELSRFKQEVSOASV 60

Db 99 LDNLLREKEVEKEHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155

Qy 61 LVGDIKTLIMDSQDKYFEATQTVYEWCGVATQLLAAAYILLFDYNEKASAKOILIKVL 120

Db 156 L-RDVTVAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSLRDVTAQ-L 205

Qy 121 DGIITKLNKAEQSLI-VSSQSFNNASGKLLAI-DSQLTN-DFSEKSSYFQSOVDKIRKEA 177
Db 206 ESQKYNDAQSLRDVSAQLESYKSTLKEIEDLKLENLTLOEKVAMAEKSVEDVQQOI 265
Qy 178 YAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIEKITSFLEKITDLK 308
Qy 227 NTVKQANKDI-----DAAKLKLTTETIAAIGETKTETTRFRFYVDYDDL 269
Db 309 NQLRQODEFRKQLEBKGRKTAENVTMTLT-----MEINKWRLLYEEL 353

RESULT 17

US-09-248-796A-18798
; Sequence 18798, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 18798

; LENGTH: 924

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-18798

Query Match 7.3%; Score 110.5; DB 4; Length 924;

Best Local Similarity 18.2%; Pred. No. 0.14;

Matches 52; Conservative 59; Mismatches 98; Indels 77; Gaps 8;

Qy 26 LYNKYLQVPIPWQTFDTIKELSR-----FKQEVSOASVILVGDITKLLMDSQDKYFEATQ 81

Db 233 LHKKYVD-----SIKDLKQDFLAFKQEAAGIINVLDLDAQLSELKQKQIDLEN 281

Qy 82 TVYEWCGVATQLLAAAYILLFDYNEKASAKOILIKVLDDGITKLNKAEQSLIIVSSQSF 141

Db 282 TKND---QIKELVSEHELQIEKISKDLTEKFKLVETQLLSKSHSHVQOFTKELIAESES 338

Qy 142 NNA-----SGKLLALDSQLTNDFSEKSSYFQSOVDKIRKEAYAGAAAGVVA 187

Db 339 QQVEEELNLTTHAKDSARILETQLSDAKESSEDYKLT-----382

Qy 188 GPFGLIISYSIAAGVVEGKLIPELKNKLSVQNFPTLSTVTVQANKDIDAANKLUTTEI 247

Db 383 -----TSEIVNDLKSQIETLKANLKL-----EEREIQNKKLQVSEL 421

Qy 248 AAIGEIKTETTRFRFYVDYDDLMLSLKKEAKKMINTCNEYQKRHG 293

Db 422 KELKELVE-----ELSNLLKLQQLHRKEIELNEQLEKLHG 459

RESULT 18

US-328-352-5793

; Sequence 5793, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: RAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 5793

```
; LENGTH: 1454
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5793

Query Match      7.3%; Score 109.5; DB 4; Length 1454;
Best Local Similarity 20.9%; Pred. No. 0.35; Indels 131; Gaps 18;
Matches 85; Conservative 54; Mismatches 137;

Qy 3 EIVADKTVEVVKNAIETADGALDLYNKYLQVWPOTFDETIKE-----LSRF 50
Db 278 QVLGASVQTLASDLADIADGALVVGIGYITRAILMKSS--AAIKBGMASLTASRQASVLA 335
Qy 51 KQESQASVLI-----VGIKTLMDSDQKYFPAET-----QTVYE 85
Db 336 QAEYAEATNALNAKAHLANVRATNAETQAK-FCGATAATRYAQQAQAAVTAATNAQTAAQ 394
Qy 86 -----WCGVAT-----QLLAAYILLFDEYNE-KKASACKDILI 117
Db 395 IKLNTATSIAGRLAKGAFGLIGWAGVATLGVMLAAYSYFNNKAEAKQKLAQAKVA 454
Qy 118 KVLDDGDTKL--NBAQKSLVSSQSFNNASGKLLALDSQLTNDF--SEKSSYFQSQVDK 172
Db 455 EKADEELKLTGNDKAKAVNDLTAF-NAQNKALEKSSRAVGSALIDIENVARGNREVEK 513
Qy 173 IRKEAYAGAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSQVQNPFTTSLNTVKQ- 231
Db 514 ISOEARTGT-----ISYTEAI-----ERLNKIKLPTDLYENLKQQAQY 552
Qy 232 ----ANKDIDAALKL-----TTEIAAIGBIKTETE-TTRFYVDY- 266
Db 553 DDNASKASLSAEKULRVEKLGCGNEAQAIAIOHQQADALGNATATEAKATKALQDYQ 612
Qy 267 ----DDLMLSLIKEA-----AKKMINTCNEYQKRHGKKTILFEVPEV 303
Db 613 AKQKDSVIDSYKSGWLDKGVTYAQAANILELQAKGMSAILSKEI 659

RESULT 19
US-09-538-092-807
; Sequence 807, Application US/09538092
; Patent No. 6753114
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-12/7,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 807
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YPR070M
US-09-538-092-807

Query Match      7.2%; Score 108.5; DB 4; Length 566;
Best Local Similarity 23.9%; Pred. No. 0.11; Indels 75; Gaps 18;
Matches 78; Conservative 43; Mismatches 131;

Qy 17 IETADGALDLYNKY-----LDQVWPOTFDETI-KELSRFKQBYSQAAVSLV- 62
Db 8 VETLSDMIELFKDYKPSITLENITRLCQTLGLSFTELSNELSRL-----STASKIIVI 63
```

```
Qy 63 -----GDIKTLMDSDQK--YFEATQTVYHWCQVATQQLAAAYILLFDEYNEKA 109
Db 64 DVDYNNKQDRIQDKVLASLNFDFYFNRDGEHEKSNILLNSLTQYKYPDLKAFHNLKF 123
Qy 110 SAQKDILIKVLDGDTKLN-EAOKSLVSSQSFNNASGKLLALDSQLTNDFSEKSSY- 165
Db 124 LYLDDAYSHIESDSTSHNNGSSDKSLDSSNASFNN-QGKL-----DLPKYFTELSHVIRQ 177
Qy 166 -FOSQV--DKIRKEAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLS 217
Db 178 CFQDNCCDFKVRTN-----LNDKFGI---YILTQG-INGKEVPLAKIYLEENKSDS 224
Qy 218 VQNFFTL--SNTVQKQKIDID--AAKLLTTEIAAIGBIKTETE-TTRFYVDY- 266
Db 225 QYRFYIYSQETKSWINEAENFSGISLUMEIVANAKESNYTDLIWFPEDFISPELII 284
Qy 267 DDLMLSLIKEAAKQKMIN--TCNEYQKR 291
Db 285 DKVTCSSNSSSSPPIIDLFNNNNYSR 311

RESULT 20
US-09-248-796A-20145
; Sequence 20145, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20145
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20145

Query Match      7.1%; Score 107.5; DB 4; Length 829;
Best Local Similarity 19.5%; Pred. No. 0.23;
Matches 66; Conservative 57; Mismatches 118; Indels 97; Gaps 12;

Qy 21 DGALDL---YNKYLDQVIPHOTFDETIKELSRFKQBYSQAAVSLVGDIKTLMDSDQKYF 77
Db 96 DGSINVTCTKTKRFTLQVLYKSF-ERLKENLVDESMTIQAYILCSWELF-----QQNYE 150
Qy 78 EATQTVYHWCQVATQQLAAAYILLFDEYNEKKAQKIDILIKVLDGDTKLNKAEQKSLVS 137
Db 151 ECWKMFHACISISFSLGHIMNQFRETN-----TKENKPKPVVDA 191
Qy 138 SQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDKTRKEAYAGA-----AAGVV 186
Db 192 TMGESDSS-----SEKDTMELEEDKNEDEELNAQQVRLYALKYLTSVI 237
Qy 187 AGPFGGLIISYSIAAGVVEGKLIPELKNK-----LKSQNFPTTSLNTVQKAKDIDAA- 239
Db 238 CSIFGRNPISVKVGMIDSTPLNQIDQKHLVLLKSESELRLNLSNLIENYMLDISFEN 297
Qy 240 ----KLKLTTEIAAI-----GEI-----KTETETTRFYVDYDDL- 270
Db 298 VMTLKLFDFDILTLEDSYDIHGKITCSLIGSEQQKSNNDANKWIGYDDYTKGSLTT 357
Qy 271 --LSLKEAAKQKMIN-----CNEYQKRHGKKTLPF 299
Db 358 KELDVLSDTITLHINSVKLLEPFVKKYETQKGKDTLSE 395

RESULT 21
```

US-09-107-532A-5150
; Sequence 5150, Application US/09107532A
; Patent No. 583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1231
; SEQUENCE DESCRIPTION: SEQ ID NO: 5150:
US-09-107-532A-5150
Query Match 7.1%; Score 106.5; DB 4; Length 1231;
Best Local Similarity 23.1%; Pred. No. 0.52;
Matches 78; Conservative 58; Mismatches 125; Indels 77; Gaps 18;
QY 6 ADKTEVVV---KNAIFADGALDLYNK-----YLDQVWPQTFPE-----TIK 45
DB 266 ADAAEEVIVSGVSAQTIETIDYLNANGKRTGFLNIHLPPFVFQKLPHTVKTIA 325
QY 46 ELSRFKQEYQASVINGDIKTLMDSQ-----DKYFEATQVYEWCGVATQLAAAYI 98
DB 326 VLDRSKFPGAGPEPLL-DVQSALYDELPAVIGRYGLGSKDV-----TPQISA--- 376
QY 99 LLFPEYNEKKAQAQDILIKVLDDGITKLENAQKSL-LVSSQSFN-----NAS 145
DB 377 -VFDELKKDPSIRKRTFIVDDVTYQSLPEKESLDLTFPQTFQAKFGFGSDGTVGAN 435
QY 146 KGLIALDSQLTNDFSEKSSYFQSQ-----VDKIR-----KEAYAGAAAGVVA--GPF 191
DB 436 KSAIKIIGDHTDKYAAQGYFYDYSKSGGLTVSHLRFGDTPIRAYLVEHADLVACHTP-A 494
QY 192 LIISYSIAAGVVEGKLIPELKNKLSVQNFTTILSNVQK--ANKDI---DAAKLKLTTE 246
DB 495 YLHSYDLVKLPGGIF--LLNTLWSDEQLTHPLPLKRYLAENNIRFTYINAMRLAQE 552

QY 247 IAAIGEKITETETTRF-----YVDYDDMLSLKKAANK 280
DB 553 VLGRRINTAMETAFFKLADIIPDE-VLPLLKEALK 589
RESULT 22
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-08-592-126-148
Query Match 7.0%; Score 105.5; DB 2; Length 1312;
Best Local Similarity 18.8%; Pred. No. 0.72;
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;
QY 10 VEVVNAIFADGALDLYNKYLDQVWPQTFPE-----ELSRFKQY 54
DB 502 MEVISLQNEKAD--LDRTLRLKQEMEQMLNHTTTRTQEMLETKADKDEQIRKIKSRH 559
QY 55 SQAASVLVGDIKTLMDSQDKYFEATQVYEWCGVATQLAAAYILLFDEYNEKKAQAQD 114
DB 560 SDELTSLG-----YFENKQLEDLHLSKSK-----EINQTR----- 591
QY 115 ILIKVLDDGITKLENAQKSLVSSQSFNNAAGKLLALDLSQLTNDFSEK-----SSYFQS 168
DB 592 -----DRLAKLN---KELASSEQNKNHINNELKREEQSLSS--YEDKLPDVCQSQDFES 640
QY 169 QVDKIRKE-----AYAGAAAGVAGPF----- 190
DB 641 DLDRLEKIEKSKQRAMLAGATA--VYSQFIQTLDENQSCCPVCQVFQTEAELOEVI 698
QY 191 -----GLI--ISYSIAAGVVEGKLIPELKNKLS 217
DB 699 SDLQSKRLAPDLKSTSELKKEKRRDEMLGLVPMRQSIID--LKEKEIPELRNKLQN 756

QY 218 VQNFPTLNTVQKANDIDAAKLUKLTETAAIGETKTTETTT-----RFVVDY 266
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQOMEL 802
QY 267 DDLMLSLLEAKA-----KMINTCNE--YQKRHGKKTLL 297
Db 803 KDVVKIAQAQAKLQIGDLDRTVQVQVQKQKQKHLDTV 842

RESULT 23
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 7.0%; Score 105.5; DB 2; Length 1312;
Best Local Similarity 18.8%; Pred. No. 0.72;
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;

QY 10 VEVVKNAIETADGALDLYNKYLQVLPWQTFDETIK-----ELSRFKQY 54
Db 502 MEVISLQNEKAD--LDRTRKLDQEMEQLNHHTTTTQMEMLTKDKADKDEQIRKIKSRH 559
QY 55 SQASVLVAGDIKTLMDSDQKYEATQTVVYEWCGVATQLLAAYILLDFDYEYNEKKSQAQKD 114
Db 560 SDELTSLG-----YFNNKQLSDMLHRSK-----EINQTR----- 591
QY 115 ILIKVLDDGDKTLKNEAQKSLVSSQSFNNASGKLLALDLSQTNDFSEK-----SSYFQS 168
Db 592 -----DLAKUN---KELASSEQNKHNINNELKRKEEQQLSS-YEDKLPDVCQSQDFES 640

QY 169 QVDKIRKE-----AVAGAAAGVVAGPF----- 190
Db 641 DDLRLKEEIEKSKQRAMLAGATA--VYSQFITQLTDENOSCCPVCORVFQTAELQOEVI 698
QY 191 -----GLI-ISYSIAAGVVEGKLIPELKNKLS 217
Db 699 SDLOSGLRLAPDKLKSTESLKKKKRDEMLGLVPMRQSIID--LKEKEIPELKNKLN 756
QY 218 VQNFPTLNTVQKANDIDAAKLUKLTETAAIGETKTTETTT-----RFVVDY 266
Db 757 V-----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQOMEL 802
QY 267 DDLMLSLLEAKA-----KMINTCNE--YQKRHGKKTLL 297
Db 803 KDVVKIAQAQAKLQIGDLDRTVQVQVQKQKQKHLDTV 842

RESULT 24
US-09-168-595-148
; Sequence 148, Application US/09168595
; Patent No. 6555666
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/168,595
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,126
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-09-168-595-148

Query Match 7.0%; Score 105.5; DB 4; Length 1312;
Best Local Similarity 18.8%; Pred. No. 0.72;
Matches 75; Conservative 58; Mismatches 96; Indels 171; Gaps 17;

QY 10 VEVVKNAIETADGALDLYNKYLQVLPWQTFDETIK-----ELSRFKQY 54
Db 502 MEVISLQNEKAD--LDRTRKLDQEMEQLNHHTTTTQMEMLTKDKADKDEQIRKIKSRH 559
QY 55 SQASVLVAGDIKTLMDSDQKYEATQTVVYEWCGVATQLLAAYILLDFDYEYNEKKSQAQKD 114

Db 560 SDELTSILG-----YFNNKQLEDWILHKSX-----EINQTR----- 591
Qy 115 ILIKVLDGTYKLEAOKSLVSSQSFNNASGKLLALDSQTLNDFSEK-----SSYFOS 168
Db 592 -----DRLAKLN--KELASSEQNKHNINELKRREQLSS-YEDKLFVDCGSDFFS 640
Qy 169 QVDKIRKE-----AYGAAAGVVGAGPF----- 190
Db 641 DLRLKEIEKSKORAMLAGATA--VYSOFITQTLTENDQSCCPVQCFVQTAEAEQEWI 698
Qy 191 -----GLI-LSYSIAAGVVEGKLIPELKNKJLS 217
Db 699 SDLOSKURLAPDKLSTESSELKKEKDERDEMLGLVPMRQSLID--LKEKEIPELKNKJLN 756
Qy 218 VQNFPTLSTNVKQANKDIDAOKLJTEIAAIGETETTT-----RYVDY 266
Db 757 V-----NRDIQRLKNDIEBQETLLGTIMPEBESAKVCLTDVTIMERFQMEI 802
Qy 267 DDLMLSLKBAK-----KWINTCNE--YQKRHGKKTIL 297
Db 803 KDKERKIAQAAKLGQIDLDRTVOOVNQEKQKQKHLDTV 842

RESULT 25
US-09-710-279-1050
; Sequence 1050, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1050
; LENGTH: 794
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1050

Query Match 6.9%; Score 104; DB 4; Length 794;
Best Local Similarity 17.0%; Pred. No. 0.47;
Matches 69; Conservative 67; Mismatches 123; Indels 148; Gaps 14;
Qy 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQVYQAASVL--VGDIK 66
Db 189 TINKIKSAVYQANNSLPKINQFADKII-----ELNKHQDDLDAYANQFRSLGKYK 238
Qy 67 TLMDSDQKYFEATQVYEWCGVATOLLAAAYILLFDEY-----NEKKSAAOK- 113
Db 239 GNVLDAAQEKLVANSSI-----PALNERAKLILDSYMPNTERILNVAANDVPAQFPRI 293
Qy 114 ----DILIKVLDGITYKLEAOKSLVSSQS-----FNNASGKLLALDSQ----- 154
Db 294 NRGVDIASEGIDAASGQLNDA--KGYLTQAKARVGDYQEAAGRAQDVNNQANQLNRQTST 352
Qy 155 -----LTN 157
Db 353 TPQSAIKSHSEKSHSSIKTVPVSGENQPVYGDNLNSDVKSMNTALTEALLSLN 412
Qy 158 DFSEKSYFQSDVKIRKEAYAGAAAGVAG-----PFGLIISYSIAAGVVEGKLIIP 209
Db 413 QTDQQAQATQODIKSLKNIAIY-----GVIASDKPSFKPEPLKNIKSRLENASKYNOQFID 467
Qy 210 -----ELKNKLSQVNFPTLSTNVKQANKDIDA-----AKLKLTTETIAAIGE 252
Db 468 ILSELEKSEHVDLSNEIKQVKEANNSINDNLKSTNQLIDALSNGSSGQLEAVNVLDLPN 527

Qy 253 IKTTETTRFYVDYDDLMMLSLKKEAAKMKMINTCNEY--QKRHGKKTIL 297
Db 528 LNKRLDTLRNVIK-----KELNRNLLAVSNEITDQLNKGQNTIL 565
RESULT 26
US-09-134-001C-4497
; Sequence 4497, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4497
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4497

Query Match 6.9%; Score 104; DB 3; Length 962;
Best Local Similarity 17.0%; Pred. No. 0.63;
Matches 69; Conservative 67; Mismatches 123; Indels 148; Gaps 14;
Qy 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQVYQAASVL--VGDIK 66
Db 199 TINKIKSAVYQANNSLPKINQFADKII-----ELNKHQDDLDAYANQFRSLGKYK 248
Qy 67 TLMDSDQKYFEATQVYEWCGVATOLLAAAYILLFDEY-----NEKKSAAOK- 113
Db 249 GNVLDAAQEKLVANSSI-----PALNERAKLILDSYMPNTERILNVAANDVPAQFPRI 303
Qy 114 ----DILIKVLDGITYKLEAOKSLVSSQS-----FNNASGKLLALDSQ----- 154
Db 304 NRGVDIASEGIDAASGQLNDA--KGYLTQAKARVGDYQEAAGRAQDVNNQANQLNRQTST 362
Qy 155 -----LTN 157
Db 363 TPQSAIKSHSEKSHSSIKTVPVSGENQPVYGDNLNSDVKSMNTALTEALLSLN 422
Qy 158 DFSEKSYFQSDVKIRKEAYAGAAAGVAG-----PFGLIISYSIAAGVVEGKLIIP 209
Db 423 QTDQQAQATQODIKSLKNIAIY-----GVIASDKPSFKPEPLKNIKSRLENASKYNOQFID 477
Qy 210 -----ELKNKLSQVNFPTLSTNVKQANKDIDA-----AKLKLTTETIAAIGE 252
Db 478 ILSELEKSEHVDLSNEIKQVKEANNSINDNLKSTNQLIDALSNGSSGQLEAVNVLDLPN 537
Qy 253 IKTTETTRFYVDYDDLMMLSLKKEAAKMKMINTCNEY--QKRHGKKTIL 297
Db 538 LNKRLDTLRNVIK-----KELNRNLLAVSNEITDQLNKGQNTIL 575

RESULT 27
US-09-489-039A-13939
; Sequence 13939, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13939
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13939

Query Match 6.9%; Score 103.5; DB 4; Length 349;
Best Local Similarity 18.9%; Pred. No. 0.15;
Matches 61; Conservative 42; Mismatches 121; Indels 99; Gaps 10;

QY 9 TVEVKAETADGALDLYNKYLDOVIPWOTFD-ETIKELSRF--KOEYSOA-----57
DB 74 SVDTDALVINGKTIIVYAEKQHHPWQAGAEVVECTGFTSAEKSOAHLQAGARK 133
QY 58 --ASVLVGDIKTLMDSDQKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI 115
DB 134 VLISAPAGEMKTIIVYVNDLTLTPDDTIISVASCTTNCLAP-----174
QY 116 LKVLDDGITKLNKAEQKSLVSSQFNNAKGLALDLSQLTNDFSEKSSYFQSQVDKIRK 175
DB 175 MAKVLQD-----AFGITVGTMTTI-----193
QY 176 EAYAGAAGVVGPRGL-----IISYSIAAGVVEGKLIPELKNKLSVQNFETT 224
DB 194 HAYTGQS-LVDGPRGKDLRASAAENVIPHTTGAAGAIGLVIPSLSGKLKHAQRVPT 252
QY 225 LSNVTQK----ANKDIDAAKLKLTEIAAAGE-----IKTETETTRFYVDYDGLM 270
DB 253 KTGSVTELVSLEKVTAEVYNQAMQQAEGNESFGYTEIEVSSDIIGHSGSIYDQ 312
QY 271 LSLLEAAKGMINTCNEYQKRHG 293
DB 313 LEIVEAGGVQLVKTVAWVDNEYG 335

RESULT 28

US-09-091-117-2
; Sequence 2, Application US/09091117
; Patent No. 6171589
; GENERAL INFORMATION:
; APPLICANT: The University of Melbourne
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
; TITLE OF INVENTION: Vaccines
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,117
; FILING DATE: 12 JUNE 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU96/00803
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN7127
; FILING DATE: 13-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WINNER, Ellen P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: +1 303 499 8080
; TELEFAX: +1 303 499 8089
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1030 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-091-117-2

Query Match 6.9%; Score 103.5; DB 3; Length 1030;
Best Local Similarity 21.2%; Pred. No. 0.77;
Matches 59; Conservative 45; Mismatches 115; Indels 59; Gaps 10;

QY 10 VEVVKAETADGALDLYNKYLDOVIP-WOTFDET-----IKLSRFKQEY-----SOA 57
DB 384 VDLVFDVITSDGDIYERKQSKDLIVPAVKNKTAALSPLIBELLTKQKTYVFDLQKH 443
QY 58 ASVLVGDIKTLMDSDQKY-FEATQTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI 116
DB 444 KGILTLLNKLFLADFQKSTPFMAQ-----VAIFTELFDEGAFDLFGFADFV 491
QY 117 IKVLDDGITKLN-----EAKSLVSSQFNNAKGLALDLSQLTNDFSEKSSYFQSQ 169
DB 492 DKIAELFLTKRTVKNGEKIETKDSLVT--SLKSLGKVAALDLDLDSYIFKNELNRS 549
QY 170 VDKIRKEAYAGAAAGVVGPRGLIISYSIAAGVVEGKLIPELKNKLSVQNFETT-L-SNT 228
DB 550 VEVAKAA-----KOTKGATDYK-----KEQAKALKKLFGHIGENT 585
QY 229 VKQAKNDIDAAKLKLTEIAAAGEIKTETETTRFYVDY 266
DB 586 LSKTNLDKITLKEVKNTEVNELESETETTLKVKKLDVEY 623

RESULT 29

US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 6.9%; Score 103.5; DB 3; Length 3696;
Best Local Similarity 22.0%; Pred. No. 5.2;
Matches 71; Conservative 49; Mismatches 144; Indels 59; Gaps 11;

QY 7 DKTVEVVKNAIETADGALDLYNKYLDOVIPWOTFDETIKLSRFKQEYSOA-ASVLVGD 65
DB 564 DDLSEQVNDIIFSNYTLASYNKYNKLERAKQTVLDEETNTFENQRYSTQTDLLHEL 623
QY 66 KTLML-----DSQDKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI 118
DB 624 QTTLINRVASREINDKAQEMTDAVYSTELTTE---EKDTLVDQIENHKNISNNIDDE 680
QY 119 VLDDGITKLNKAEQKSLVSS-----QSFNNAKGLALDLSQLTNDFSEKSSYFQ 167
DB 681 LTDDGVERVKEAGLHTLESPTHPVKPNARQVNNRA-----DQDKTLIRNHEATTE 734
QY 168 SQVDKIRK-EAYAGAAAGVVGPRGLIISYSIAAGVVEGKLIPELKNKLSVQNFETT-L 226
DB 735 EQNEAIRQVEAHSSDAIA-----KIGAEETDTTNEARDNGTKLIA---TDVP 779

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4284

Query Match
Best Local Similarity 17.5%; Pred. No. 0.61;
Matches 58; Conservative 54; Mismatches 130; Indels 90; Gaps 9;

QY 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQ--EYSQAASVVLVDGIK 66
Db 220 TAHIIQSMADKADTEIIIPQITLFGLEPFDEVIKEAESFKEPAEKDREGRVLDRLDA 279
QY 67 TLLMDSQD-KYFEATQTVYEWCG-----VATQLLAAYILLFDEYNEKKAQAQKDI---- 115
Db 280 LVTIGEDARDDDAVYAEKPGGGYRVVVAIDVSHVYRLDSALNEEAEREGTSVYPPH 339
QY 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNASGKL-----LALDSQLTNDFSEKS 163
Db 340 FVLPMLEALSNGLSLNPHVDRLCMVCDLKLSTRTGVTGYEFYFPAVNHKSARLTYYTQVG 399
QY 164 SYFQSQVDKIRKEAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNLKSQNFFT 223
Db 400 QYFEGATDAIPKDR-----DIHKSILNTLFQLYQ 427
QY 224 TILSNVTYKQANKIDAAKLTTEIAAIGETETETTFYDYDDL-----MLSLKEAA 278
Db 428 ILKN-----LRVDRHAMEFETIETMTFDELGGIKELIPRTRNDA 467
QY 279 KQWINTC-----NEYQKRGHKKTLPEVPE 302
Db 468 HKLIEECMLLANVAAYALEHDIPMLYRVHE 499

RESULT 32
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5178

Query Match
Best Local Similarity 19.5%; Pred. No. 0.84;
Matches 65; Conservative 63; Mismatches 130; Indels 76; Gaps 13;

QY 11 EVVKNAIETADGALDLYNKYLDQ-----VIPWQTFDETIKELSRFKQ----- 52
Db 107 ELVDNQFELRESKVGNGQFVQLLGVNAEQFRQLFILPQGEFKFLQSNKDKQSILRT 166
QY 53 -----EYSQAASVVLVDGIKTLMSQDKYFEATQTVYEWCGVAT---QLLAAY----- 97
Db 167 LFNSERFDEIRLLVENVYKQKVIENRY---TQIENLWMDIDITFNDELALYKELESSQ 223
QY 98 -----ILLFDEYNE-----KKASAQKDIILIKVLDDGITKLINEAQKSLVSSQSFNASGK 147
Db 224 TDKMIEKFPQFNDYGGCKILKSPFEAKNKITKELDD---LNHKKYKVNVELSENTKKLAE 279
QY 148 LLALDSQLTNDFSEKSYFSQSDVKIRKEAYAGAAAGVAGPFGGLIISYS 197
Db 280 KIKFD-----DLKKEQY---IDKLQELKMIQESKVLITVFTRLQSLKDKDELVSILH 330
QY 198 IAAGVVEGKLIPELKNLKSQNFFTLNTYKQANKIDAAKLTTEIAAIGELK--- 254
Db 331 EQSKLNETNYHNEIKGFKQKLEH--LSTRENEITQFNQYLEKNGQVFNQDLDKLISSYQKXP 389
QY 255 -TETETRFVYDDMLSLKEAAKQWINTCNE 287
Db 390 VIEEIKRLSEYNDLITK--KEELYEMNNKNK 421

RESULT 31
US-09-328-352-4284
; Sequence 4284, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04

```

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4284
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4284

Query Match
Best Local Similarity 17.5%; Pred. No. 0.61;
Matches 58; Conservative 54; Mismatches 130; Indels 90; Gaps 9;

QY 9 TVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKQ--EYSQAASVVLVDGIK 66
Db 220 TAHIIQSMADKADTEIIIPQITLFGLEPFDEVIKEAESFKEPAEKDREGRVLDRLDA 279
QY 67 TLLMDSQD-KYFEATQTVYEWCG-----VATQLLAAYILLFDEYNEKKAQAQKDI---- 115
Db 280 LVTIGEDARDDDAVYAEKPGGGYRVVVAIDVSHVYRLDSALNEEAEREGTSVYPPH 339
QY 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNASGKL-----LALDSQLTNDFSEKS 163
Db 340 FVLPMLEALSNGLSLNPHVDRLCMVCDLKLSTRTGVTGYEFYFPAVNHKSARLTYYTQVG 399
QY 164 SYFQSQVDKIRKEAYAGAAAGVAGPFGGLIISYSIAAGVVEGKLIPELKNLKSQNFFT 223
Db 400 QYFEGATDAIPKDR-----DIHKSILNTLFQLYQ 427
QY 224 TILSNVTYKQANKIDAAKLTTEIAAIGETETETTFYDYDDL-----MLSLKEAA 278
Db 428 ILKN-----LRVDRHAMEFETIETMTFDELGGIKELIPRTRNDA 467
QY 279 KQWINTC-----NEYQKRGHKKTLPEVPE 302
Db 468 HKLIEECMLLANVAAYALEHDIPMLYRVHE 499

RESULT 32
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5178

Query Match
Best Local Similarity 19.5%; Pred. No. 0.84;
Matches 65; Conservative 63; Mismatches 130; Indels 76; Gaps 13;

QY 11 EVVKNAIETADGALDLYNKYLDQ-----VIPWQTFDETIKELSRFKQ----- 52
Db 107 ELVDNQFELRESKVGNGQFVQLLGVNAEQFRQLFILPQGEFKFLQSNKDKQSILRT 174
QY 53 -----EYSQAASVVLVDGIKTLMSQDKYFEATQTVYEWCGVAT---QLLAAY----- 97
Db 175 LFNSERFDEIRLLVENVYKQKVIENRY---TQIENLWMDIDITFNDELALYKELESSQ 231
QY 98 -----ILLFDEYNE-----KKASAQKDIILIKVLDDGITKLINEAQKSLVSSQSFNASGK 147
Db 232 TDKMIEKFPQFNDYGGCKILKSPFEAKNKITKELDD---LNHKKYKVNVELSENTKKLAE 287

```

Qy 148 LLAIDSLQTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFF-----GLIISYS 197
Db 288 KIKED-----DLKKEQNY-----IDKLOELMIQESKVLITYFTRLOSLKKDKDELVS LH 338
Qy 198 IAAAGVGEKLPKLNKLSVQNFPTTSLNTVQKNDIDAOKLKTTEIAAIGEIK-- 254
Db 339 EOSKLNETHYNEIKGFKQLEH--LSTRENIITQNVLEKNQVFFNQLOKLISSYQOKP 397
Qy 255 -TETTRFYVDYDMLSLKEAAKMINTCNE 287
Db 398 VIEBEIKRLSEYNDLITK--KEELTKEMNNKNK 429

RESULT 33

US-09-248-796A-20613
; Sequence 20613, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20613
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20613

Query Match 6.8%; Score 102; DB 4; Length 722;
Best Local Similarity 20.0%; Pred. No. 0.63; Mismatches 99; Indels 54; Gaps 9;
Matches 51; Conservative 51;
Qy 38 QTFDETIKELSRFQKQYSQAASVLVGDITKLLMDSQDKYFEATQ-----TVYEWCGVATQ 92
Db 211 QIQYDITKLSKSTREELNGSKTEIL--RLKALLRESEBELYQVKENYKTSVHDYEQDLAQ 269
Qy 93 LLAAYILLFDEYNKKSAAQKDIILIKVLDGITKLNKLSVQSFNNASGKLALD 152
Db 270 LKVKHETLLS--RNKDINESLEIYKGRSDEYKKELEAESAIAISKRHEQATKEMKESR 327
Qy 153 SQ-----LTDNFEKSSYFQSQVDKIRKEAYAGAAAGVAGPFFGLIISYSIA 199
Db 328 SQQLLVREELRTTQILKDFRIKVENLEATIE-----KNHOLD 366
Qy 200 AGVVEGKLIPK--LKNKLSVQNF--FFTTLSNTVQKNDIDAOKLKTTEIAAIGEIKTE 256
Db 367 ANKEEIKQIQKLNKLNKFNENKELNEKLEKKEIKNLNRLD-----FKTDI-----ETKLI 417
Qy 257 TETTRFYVDYDML 271
Db 418 KENKQLQDYEDVLL 432

RESULT 34

US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 6.8%; Score 102; DB 3; Length 10182;
Best Local Similarity 20.7%; Pred. No. 33;
Matches 65; Conservative 47; Mismatches 104; Indels 98; Gaps 12;
Qy 9 TVEVVKNAIETADGALDLYNKYLDQVIPWOTFDETIKELSRFQKQYSQAASVLVGDITKL 68
Db 4407 TVEQVKESVANA-----QQVI-----QDLQARTSLVDP-KTQ 4438
Qy 69 LMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNKKSAAQKDIILIKVLDGITKL 128
Db 4439 LQEAQNRLSNSINOQDTDTGMDQSL-----NNYNDKLAKARQNL-----EKISKVL 4485
Qy 129 EAQKSLIVSSQSFNNASGKLALD---SOLT-----NDFSEKSSYFQSQVDKIRKEAY 178
Db 4486 GGQPTVAEIRQNTDEANAHKQALDTARSQTLNREPYNHNNESHNNNAQKDNFK---- 4541
Qy 179 AGAAAGVAGPFFGLIISYSIAAGVVEGKLIPKLNKLSVQNFPTTSLNTV-----KQA 232
Db 4542 ----AQVNSAP-----NHNLTETIKRKAADTLNQSMTALSESIAQYENQKQ 4583
Qy 233 NKDIDAOKL---LTTEIAAIGEIKTETETTRFYVDYDD-----LMLS 273
Db 4584 ENYLDASNNKRDYDNVNAAKGILNQTSPTMSADVIDOKAEDVKKTKTALDGNQRLEV 4643
Qy 274 LKEAAKKMINTCNE 287
Db 4644 AKQOALNHLNTLND 4657

RESULT 35

US-09-492-709A-302
; Sequence 302, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITPA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 302
; LENGTH: 2383
; TYPE: PRT
; ORGANISM: E. Coli
US-09-492-709A-302

Query Match 6.7%; Score 101; DB 4; Length 2383;
Best Local Similarity 22.9%; Pred. No. 4.7;
Matches 61; Conservative 51; Mismatches 102; Indels 52; Gaps 11;
Qy 41 DETIKELSRFQKQYSQAASV--LVGDIKTLMDSDQKYFEATQTVY--EWCGVATQLL--- 94
Db 685 DENDKPKVEKQKQQLNNAVSIDNVKPGVTTDMDKADGVYKATYATYKSGSLTAKLLMQN 744

95 QY -----AAVILLFDEYNEKA-----SAQKILIKVLDGKITKLEAKSL 134
 745 DB WNEHLHTAGFIIDANPQSAKIATLSASNGVLANENAAVTSVNVADEGSNFINDHVTTF 804
 135 QY LV-----SSQSFNAGSKLLALDSQITNDFSEKSYFQSDVKIRKAYAGAAAGVAGPFG 191
 805 DB AVLSGSATSFNNQNTAKTDVNGLAT--FDLKSQEDNTVEVLE-----NGVKQ---T 853
 192 QY LIISVSTAAGVVEGLIPELKNKLKSVQNFPTLSTNTVQKANK-----DIDAACL 241
 854 DB LIVSFVGDSSSTAQVDL-QSKNEVVADGNSVITMTATVRDAKGNLLNDVMVTVNNVNSAEA 912
 242 QY KLT-TEIAAIGBIKTEITETTFYDY 266
 913 DB KLSQTEVNSHDGIATATLTLKNGDY 938

RESULT 36
 US-08-434-000A-4
 ; Sequence 4, Application US/08434000A
 ; Patent No. 6046037
 ; GENERAL INFORMATION:
 ; APPLICANT: ANDREW C. HIATT, JULIAN
 ; APPLICANT: K.-C. MA, THOMAS LEHNER
 ; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
 ; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; SUITE: Suite 4700
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/434,000A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA: including application 1
 ; PRIOR APPLICATION DATA: described below:
 ; APPLICATION NUMBER: 08/367,395
 ; FILING DATE: 12/30/94
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Guise, Jeffrey W.
 ; REGISTRATION NUMBER: 34,613
 ; REFERENCE/DOCKET NUMBER: 212/127
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 552-8400
 ; TELEFAX: (619) 552-0159
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 746 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; TOPOLOGY: DESCRIPTION: Human Polymunoglobulin Receptor
 ; US-08-434-000A-4

Query Match 6.7%; Score 100.5; DB 3; Length 746;
 Best Local Similarity 22.6%; Pred. No. 0.92;
 Matches 45; Conservative 41; Mismatches 70; Indels 43; Gaps 11;
 QY 86 WCGVAT-----QLLAAYILLFDEYNEKASAKQDILI-----KVLDGKITKLEAKSL 131

525 DB WCGVKQGHFYGETAAVTVAV-----EERKAAGSRDVSIAKADAAPDEKVLDSG---FREIE 577
 132 QY KSLIVSSQSFNAGSKLLALDSQITNDFSEKSYFQSDVKIRKAYAGAAAGVAG--P 189
 578 DB NKATQDRLP--AEKAVA-----DTRQADGSRASVDSGSSEGGSSRALVSTLP 628
 190 QY FGLIISY-STAAGVVEGLIPELKNKLK-SVQNFPTLSTNTVQK-----ANKDIDAACL 241
 629 DB LGLVLAVGAVAGVARAR--HRKNVDRVIRSRYTIDISMSDFENSREFGANDNMGASSI 685
 242 QY KLTTEIAAIGBIKTEITETTT 260
 686 DB TQETSLGGKEEFVATTEST 704

RESULT 37
 US-09-312-157-4
 ; Sequence 4, Application US/09312157
 ; Patent No. 6303341
 ; GENERAL INFORMATION:
 ; APPLICANT: ANDREW C. HIATT, JULIAN
 ; APPLICANT: K.-C. MA, THOMAS LEHNER
 ; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
 ; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; SUITE: Suite 4700
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/312,157
 ; FILING DATE: 14-May-1999
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/434,000
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Guise, Jeffrey W.
 ; REGISTRATION NUMBER: 34,613
 ; REFERENCE/DOCKET NUMBER: 212/127
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 552-8400
 ; TELEFAX: (619) 552-0159
 ; TELEX: 67-351
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 746 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; TOPOLOGY: DESCRIPTION: Human Polymunoglobulin Receptor
 ; US-09-312-157-4

Query Match 6.7%; Score 100.5; DB 3; Length 746;
 Best Local Similarity 22.6%; Pred. No. 0.92;
 Matches 45; Conservative 41; Mismatches 70; Indels 43; Gaps 11;
 QY 86 WCGVAT-----QLLAAYILLFDEYNEKASAKQDILI-----KVLDGKITKLEAKSL 131
 DB 525 WCGVKQGHFYGETAAVTVAV-----EERKAAGSRDVSIAKADAAPDEKVLDSG---FREIE 577

Qy 132 KSLVSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDKIRKEAYAGAAAGVAG--P 189
Db 578 NKAIQDPLRF--AEKAVA-----DTRQADGSRASVDSSGSEEGGSRALVSTLVP 628
Qy 190 FGLIISY-SIAAGVVEKLIPELKNKLLK-SVONPFTTSLNTVKQ-----ANKDIDAACL 241
Db 629 LGLVLAVGAVGVARAR---HRKNVDRVRSIRYRTDMSDFNSREFGANDMGASSI 685
Qy 242 KLTTEIAAIGEIKTETETT 260
Db 686 TQETSLGGKEEFVATTEST 704

RESULT 38

US-09-315-793-52
; Sequence 52, Application US/09315793
; Patent No. 6221597
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; FILE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-048
; CURRENT APPLICATION NUMBER: US/09/315,793
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 52
; LENGTH: 1093
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-315-793-52

Query Match 6.7%; Score 100.5; DB 3; Length 1093;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 63; Conservative 47; Mismatches 88; Indels 93; Gaps 13;

Qy 24 LDLYNKYLQDVIPWQTFDETIKELSRFKQBYSOAASVVLVDGDKT---LLMDSQDKYFEAT 80
Db 664 INLKNEYNDRK---STLDALSNQSGYRHELSELAS-KNDDINREAHQNLNEIRKYYTMRK 719
Qy 81 QTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI--LIKVLDDGI-----TKL 127
Db 720 STIE-----TLREKLDQLKREARKDVSKIKDIDDQIQOQLLLKORHLLSKM 765
Qy 128 NEAQKSL-----LVSSQ-----SFNNASGKLLALDSQLTNDSEKSSY----- 165
Db 766 ASSMKSLKNCKQKELISTQILQFEAQNMDVSNVDVIGFFNEREADLKSQYEDKKKFKVKEMR 825
Qy 166 ----FQSQVDKIR-----KEAYAGAAAGV-VAGPFGI-----IISYSIA 199
Db 826 DTPEFQSWREIRSYDQTKELKNKVAEKVEEGNFNLSFVQDVLDKLESEIAMVNHDES 885
Qy 200 AGVVEGKLIPELKNKLSVQNFFTTLSNTVKQAKDIDAAKLKTTEIAAI 250
Db 886 AVTILQVTAELRE-----LEHTVPQSKDLETIKAKLKEDHAVL 925

RESULT 39

US-09-538-092-701
; Sequence 701, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformatter Version 0.9
; SEQ ID NO 701
; LENGTH: 1093
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YOL034W
US-09-538-092-701

Query Match 6.7%; Score 100.5; DB 4; Length 1093;
Best Local Similarity 21.8%; Pred. No. 1.6;
Matches 63; Conservative 47; Mismatches 88; Indels 93; Gaps 13;

Qy 24 LDLYNKYLQDVIPWQTFDETIKELSRFKQBYSOAASVVLVDGDKT---LLMDSQDKYFEAT 80
Db 664 INLKNEYNDRK---STLDALSNQSGYRHELSELAS-KNDDINREAHQNLNEIRKYYTMRK 719
Qy 81 QTVYEWCGVATQLLAAYILLFDEYNEKKAQAQDI--LIKVLDDGI-----TKL 127
Db 720 STIE-----TLREKLDQLKREARKDVSKIKDIDDQIQOQLLLKORHLLSKM 765
Qy 128 NEAQKSL-----LVSSQ-----SFNNASGKLLALDSQLTNDSEKSSY----- 165
Db 766 ASSMKSLKNCKQKELISTQILQFEAQNMDVSNVDVIGFFNEREADLKSQYEDKKKFKVKEMR 825
Qy 166 ----FQSQVDKIR-----KEAYAGAAAGV-VAGPFGI-----IISYSIA 199
Db 826 DTPEFQSWREIRSYDQTKELKNKVAEKVEEGNFNLSFVQDVLDKLESEIAMVNHDES 885
Qy 200 AGVVEGKLIPELKNKLSVQNFFTTLSNTVKQAKDIDAAKLKTTEIAAI 250
Db 886 AVTILQVTAELRE-----LEHTVPQSKDLETIKAKLKEDHAVL 925

RESULT 40

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949

DB	2738	EKNRLAGELQLLEIKSSKDQLKELTLENSELKSLDKCHKDQVEKGVREIEAYQL	2797
QY	100	LFDEYNEKKASA-----QKDILIKVLDGKITKLENAQKS-----LLVSSQSFNNAS	145
DB	2798	RLHE-AEKHQALLDNTNKQYEVEIQTYREKLTSEKCLSSQKLEIDLLKSSKEELNNS-	2855
QY	146	GKLLALDSQLTNDFSEKSSYFQSOVDKIRKEAYAGAAAGVAGPFGLIISYSIAAGVVEG	205
DB	2856	---LKATTQILEELKTKMDNLKYVNLKENER-----AQGKMKLLIK-SCKQLEEEK	2905
QY	206	KLIPELKNKLKSVQNFFTLNNTVKQANKDIDAAKLKTTEIAAIGE-IKTETETTRFYV	264
DB	2906	EILQKELSQAQAQEKQKT--GTVMDTKVD-----ELTTEIKELKETLEEKTKADEYL	2957
QY	265	D-YDDMLSLK-EAAKGMINT-----CNEYQKRGHK	294
DB	2958	DKYCSLLISHEKLEKAKEMLETQVAHLCSQQSKQDSR	2994

RESULT 42

US-08-353-700-1

; Sequence 1, Application US/08353700

; Patent No. 5599919

; GENERAL INFORMATION:

; APPLICANT: YEN, TIMOTHY J.

; APPLICANT: RATNER, JEROME B.

; TITLE OF INVENTION: NUCLEIC ACID ENCODING A

; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,

; TITLE OF INVENTION: AND METHODS OF USE

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN

; STREET: 1601 MARKET STREET, SUITE 720

; CITY: PHILADELPHIA

; STATE: PA

; COUNTRY: USA

; ZIP: 19103-2307

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/353,700

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: REED, JANET E.

; REGISTRATION NUMBER: 36,252

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 563-4100

; TELEFAX: (215) 563-4044

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3248 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: HUMAN

US-08-353-700-1

Query Match

Best Local Similarity

Matches

6.7%; Score 100.5; DB 1; Length 3248;

22.0%; Pred. No. 8.3;

74; Conservative

63; Mismatches 129; Indels 71; Gaps 16;

QY

9

TVEVVKNAIETADGALDI-----YNKYLDQVIPWQ-TFDETIKELSRFKQYSQ

DB

2678

TLEVLQSSYKNLELELTCKDKMSFVEKVNKMTAKETELQREHMAQKTAELQBELSG

2737

DB	2738	EKNRLAGELQLLEIKSSKDQLKELTLENSELKSLDKCHKDQVEKGVREIEAYQL	2797
QY	100	LFDEYNEKKASA-----QKDILIKVLDGKITKLENAQKS-----LLVSSQSFNNAS	145
DB	2798	RLHE-AEKHQALLDNTNKQYEVEIQTYREKLTSEKCLSSQKLEIDLLKSSKEELNNS-	2855
QY	146	GKLLALDSQLTNDFSEKSSYFQSOVDKIRKEAYAGAAAGVAGPFGLIISYSIAAGVVEG	205
DB	2856	---LKATTQILEELKTKMDNLKYVNLKENER-----AQGKMKLLIK-SCKQLEEEK	2905
QY	206	KLIPELKNKLKSVQNFFTLNNTVKQANKDIDAAKLKTTEIAAIGE-IKTETETTRFYV	264
DB	2906	EILQKELSQAQAQEKQKT--GTVMDTKVD-----ELTTEIKELKETLEEKTKADEYL	2957
QY	265	D-YDDMLSLK-EAAKGMINT-----CNEYQKRGHK	294
DB	2958	DKYCSLLISHEKLEKAKEMLETQVAHLCSQQSKQDSR	2994

RESULT 41

US-09-538-092-1154

; Sequence 1154, Application US/09538092

; Patent No. 6753314

; GENERAL INFORMATION:

; APPLICANT: Glot, Loic

; APPLICANT: Mansfield, Traci A.

; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same

; FILE REFERENCE: 15966-542

; CURRENT APPLICATION NUMBER: US/09/538,092

; CURRENT FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/127,352

; PRIOR FILING DATE: 1999-04-01

; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01

; SOFTWARE: CuraPatSeqFormatter Version 0.9

; NUMBER OF SEQ ID NOS: 1387

; SEQ ID NO 1154

; LENGTH: 3210

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (0)...(0)

; OTHER INFORMATION: Polypeptide Accession Number P49454

US-09-538-092-1154

Query Match

Best Local Similarity

Matches

6.7%; Score 100.5; DB 4; Length 3210;

22.0%; Pred. No. 8.2;

74; Conservative

63; Mismatches 129; Indels 71; Gaps 16;

QY

9

TVEVVKNAIETADGALDI-----YNKYLDQVIPWQ-TFDETIKELSRFKQYSQ

DB

2678

TLEVLQSSYKNLELELTCKDKMSFVEKVNKMTAKETELQREHMAQKTAELQBELSG

2737

QY

57

AASVLVGDIKTLMD---SQKYFEAT-----QTVYEWCGVATQLLAAYIL

99

[illegible]

QY 175 -----KEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNK 214
DB 319 VYEMQDKVNSNKEBIOKET-----IIELNTKIIEBEKKTLEKOK 363
QY 215 LKSVQNFPTLSNTVKQAKNDIDAALKLTTEIAAIGEIKTETTRFYVDYDDLMLSL 274
DB 364 LTTADKLLGELQEQIVQKQEIKNKLELT-----NSKQKQSS----- 403
QY 275 KEAAKMMINTCNEYQKRGKKTLP 299
DB 404 -EIKQLMGTVLELQKRNKDSQPE 427

RESULT 48

US-09-447-497-15
; Sequence 15, Application US/09447497
; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.
; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Sustin, Santos A.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; CURRENT FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human apoptosis-inducing factor (hAIF isoform #2);
; OTHER INFORMATION: hAIF-exon-skip-Gold; mature polypeptide
US-09-447-497-15

Query Match 6.6%; Score 99; DB 4; Length 526;
Best Local Similarity 21.8%; Pred. No. 0.76;
Matches 60; Conservative 46; Mismatches 93; Indels 76; Gaps 13;

QY 37 WQTFDETIKELSRFKQ-----EYSQAAS-----VLVGDIKTLMD 72
DB 93 WFSDDPNVTKTLRFKQWNGKERSIYFQPPSYSAQDLPHIENGVAULTGK-KVQLDV 151
QY 73 QD---KYFEATQTVYEWCGVATQLLAAYILLFDEYNEKASQAQ---KDILIKVLDD--GI 124
DB 152 RDNMVKLNDGSGQIYKELCLATGTGTPRSLAID-----RAGAEVKSRITLFRKIGDPRSL 206
QY 125 TKLNEAKSLVSSQSFNNA-----SGKLLALDSQLTNDFSEKSS-----YFQS-QV 170
DB 207 EKISREVKSTIIGGGFLGSELACALGRKARALGTEVIQLPFEKGNMGKILPEYLSNWTM 266
QY 171 DKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTLSNTVK 230
DB 267 EKVRRE-----GVKMPNAIVQS-----VGVSSGKLLIKLKDGRK-----V 302
QY 231 QANKDIDAALKLTTEIAAIGEIKTETTRFYVD 265
DB 303 ETDHVAAGVLEPNVELAKTGGLSDSDFGGFRVN 337

RESULT 49

US-09-447-497-14
; Sequence 14, Application US/09447497
; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.

; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Sustin, Santos A.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; CURRENT FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-447-497-14

Query Match 6.6%; Score 99; DB 4; Length 553;
Best Local Similarity 21.8%; Pred. No. 0.82;
Matches 60; Conservative 46; Mismatches 93; Indels 76; Gaps 13;

QY 37 WQTFDETIKELSRFKQ-----EYSQAAS-----VLVGDIKTLMD 72
DB 120 WFSDDPNVTKTLRFKQWNGKERSIYFQPPSYSAQDLPHIENGVAULTGK-KVQLDV 178
QY 73 QD---KYFEATQTVYEWCGVATQLLAAYILLFDEYNEKASQAQ---KDILIKVLDD--GI 124
DB 179 RDNMVKLNDGSGQIYKELCLATGTGTPRSLAID-----RAGAEVKSRITLFRKIGDPRSL 233
QY 125 TKLNEAKSLVSSQSFNNA-----SGKLLALDSQLTNDFSEKSS-----YFQS-QV 170
DB 234 EKISREVKSTIIGGGFLGSELACALGRKARALGTEVIQLPFEKGNMGKILPEYLSNWTM 293
QY 171 DKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLIPELKNKLSQVQNFPTLSNTVK 230
DB 294 EKVRRE-----GVKMPNAIVQS-----VGVSSGKLLIKLKDGRK-----V 329
QY 231 QANKDIDAALKLTTEIAAIGEIKTETTRFYVD 265
DB 330 ETDHVAAGVLEPNVELAKTGGLSDSDFGGFRVN 364

RESULT 50

US-09-447-497-12
; Sequence 12, Application US/09447497
; Patent No. 6773911
; GENERAL INFORMATION:
; APPLICANT: Penninger, Josef M.
; APPLICANT: Kroemer, Guido P.
; APPLICANT: Siderovski, David P.
; APPLICANT: Zamzami, Naoufal
; APPLICANT: Sustin, Santos A.
; APPLICANT: Snow, Bryan E.L.
; TITLE OF INVENTION: APOPTOSIS-INDUCING FACTOR
; FILE REFERENCE: 01017/36780
; CURRENT APPLICATION NUMBER: US/09/447,497
; CURRENT FILING DATE: 1999-11-23
; EARLIER APPLICATION NUMBER: 60/109,595
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human apoptosis-inducing factor (hAIF isoform #1)
; OTHER INFORMATION: hAIF-alt-exon-Gold; mature polypeptide
US-09-447-497-12

Query Match 6.6%; Score 99; DB 4; Length 609;

Best Local Similarity 21.8%; Pred. No. 0.94;
Matches 60; Conservative 46; Mismatches 93; Indels 76; Gaps 13;

[illegible]

Search completed: January 5, 2005, 10:59:05
Job time : 20.9474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:52:17 ; Search time 17.4424 Seconds
(without alignments)
1671.423 Million cell updates/sec

Title: US-09-993-292B-28
Perfect score: 1508
Sequence: 1 MTEIVADKTVEVVKNAIETA.....TCNEYQKSHCKKTLFVPEV 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : PIR 79:**

1: Piri:**

2: Piri:**

3: Piri:**

4: Piri:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1508	100.0	305	2 C64864	hemolysin E - Esch
2	1493	99.0	305	2 E90838	hemolysin E [impor
3	1493	99.0	305	2 E85696	probable pore form
4	1400	92.8	305	2 AE0673	haemolysin HlyE [i
5	133.5	8.9	584	2 S75986	hypothetical prote
6	126	8.4	652	2 B59102	hypothetical prote
7	122	8.1	1496	2 T05634	hypothetical prote
8	120	8.0	1039	2 S18199	myosin heavy chain
9	117.5	7.8	927	2 AH1369	transmembrane prot
10	117.5	7.8	1098	2 B70232	hypothetical prote
11	116.5	7.7	927	2 AG1739	transmembrane prot
12	115.5	7.7	664	2 A97222	membrane associate
13	115	7.6	1127	2 T28317	ORF MSV156 hypothe
14	115	7.6	1964	2 A59282	nonmuscle myosin I
15	115	7.6	2017	1 A36014	myosin heavy chain
16	115	7.6	2057	2 S61477	myosin II heavy ch
17	114.5	7.6	1269	2 E84730	probable myosin he
18	114	7.6	796	2 B84800	probable alpha-car
19	114	7.6	1738	2 T14867	interactin - slime
20	114	7.6	2401	2 T28676	rophry protein -
21	113.5	7.5	1024	2 S10056	hemolysin A - Esch
22	113	7.5	622	2 T22716	hypothetical prote
23	113	7.5	955	1 A35244	leukotoxin A - Pas
24	112.5	7.5	1679	2 S48385	hypothetical prote
25	112	7.4	295	2 G97827	hypothetical prote
26	112	7.4	631	2 JC4298	hyaluronan recepto
27	112	7.4	1875	2 S38173	myosin-like protei
28	111.5	7.4	821	2 S67087	hypothetical prote
29	111.5	7.4	2139	2 T18296	myosin heavy chain

ALIGNMENTS

RESULT 1

C64864

hemolysin E - Escherichia coli (strain K-12)

N;Alternate names: hemolysin-inducing protein

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C;Accession: C64864

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: C64864

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-305 <BLAT>

A;Cross-references: GB:AE000216; GB:U00096; NID:gl787417; PIDN:AAC74266.1; PID:gl787430;

A;Experimental source: strain K-12, substrain MGL55

C;Genetics:

A;Gene: hlyE; hpr

C;Function:

A;Description: hemolytic activity

A;Note: pore formation

C;Superfamily: Escherichia coli hemolysin E

C;Keywords: cytotoxic; cytotoxin; hemolysis; transmembrane protein

F;181-197/Domain: transmembrane #status predicted <TM>

F;123/Active site: Asp #status predicted

Query Match 100.0%; Score 1508; DB 2; Length 305;

Best Local Similarity 100.0%; Pred. No. 2.4e-87;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKLSRFKQEYSQAASV 60

Db 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVTPQTFDETIKLSRFKQEYSQAASV 62

Qy 61 LVGDITKLMDSDQKQFEATQTVYVWCGVATQLLAAYILLFDYNEKKAQAQDKILIKVL 120

Db 63 LVGDITKLMDSDQKQFEATQTVYVWCGVATQLLAAYILLFDYNEKKAQAQDKILIKVL 122

Qy 121 DDGITKLNEAKSLLYSQSFNNAGSKLLALDLSQLTNDFSEKSYFQSOVDKIRKEAYAG 180

Db 123 DDGITKLNEAKSLLYSQSFNNAGSKLLALDLSQLTNDFSEKSYFQSOVDKIRKEAYAG 182

Qy 181 AAAGVAGPPGLIISYSIAAGVVEGKLIPELKNKLSVQNFFTTLSNTVQANKDIDAAK 240

Db 183 AAAGVAGPPGLIISYSIAAGVVEGKLIPELKNKLSVQNFFTTLSNTVQANKDIDAAK 242

R.; Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, C.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, A.C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: ABL077; MUID:21537279; PMID:11679669
A:Accession: AHI369
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-927 cGLA>
A:Cross-references: UNIPROT:Q8Y4S2; GB:NC_003210; PID:NCAD00438.1; PID:G16411848; GSPDB:1-927
C:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2360

Query Match	7.8%;	Score 117.5;	DB 2;	Length 927;
Best Local Similarity	21.1%;	Pred. No. 9.7;		
Matches	73;	Conservative	33;	Mismatches 141; Indels 99; Gaps 11;
QY	3	EIVADQTVVKKNAIETADGALDLYNKYLDQVPIWQTFDETI----	KELSRFKQEYSQAA	58
DB	166	EAIFDKIKESGSGFQAADGS-----GKIKDGLVKSGEKNKTISTNLKTLADSSLTFDKGA	221	
QY	59	SVL-----VGDIK-----TLMDSQDKYF	77	
DB	222	NTEVGLKTYTDGVTAAAGDKLNAGVSTLAAGVGLKDGVAALDGGATKLASGVSTYT	281	
QY	78	EATQTVVWCGVATQALLAYILLFDYNEKKASAKXDIILIKVLDDGITKLNKAQSKLLVS	137	
DB	282	SGVDTL---AGGINQAVTGSTALSDGLNKWGS-----VPTLASGITQLNNGQKSLATG	332	
QY	138	SQSFNNASGKLLA-----LDSQLTN-----DFSEKSSYFQSQVDKIRKEAYAGAAAGVWA	187	
DB	333	LDSLVDSGNKLSAGLKELDGNLTDKOGKIAQLKQGMNDLQQGDIDQLNQSVNGEDAA----	388	
QY	188	GPFCLLIISYSIAAGVWEGKLIPELKNKLSVQNFFTTLSNTVVKQAKNDIDAAKLKUTTEI	247	
DB	389	-----IAKQIATLQKLSLDLQNLGTLFIKS---NANFDAEAIKSKINATA	429	
QY	248	AAIGEIKETETTRFYVDYDDMLMLSLLEKAAKKVINTCNEYQKRHG	293	
DB	429	-----IAIQAQDLKKTOKSAQVATVTEOLOSG	467	

RESULT 10
B70232
hypothetical protein BBG10 - Lyme disease spirochete plasmid G/lp28-2
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: B70232
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, J.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: B70232
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1098 <L>
A:Cross-references: UNIPROT:O50733; GB:AE000786; NID:g2690008; PID:g2690008; Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 7.8%; Score 117.5; DB 2; Length 1098;
Best Local Similarity 20.2%; Pred. No. 12;
Matches 78; Conservative 63; Mismatches 123; Indels 123; Gaps 17;
QV 10 VEVVKNALETDGAALDLYNKY-----LDQVIPQGTDETIKELSR 49

```

YY      110 SAQKDI.LKVLDGGITTKNEAQSKLLVSSQSFNNAAGKKLALDSQTNDPSEKSIFYQ--- 167
DBB     :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|
        264 ETQVEFEAKYEDAFACQDARTQL-----DDLGRKDWEVAELRQTLMSMKDAYFKEM 316
QQ      168 ----SQVDKIRKEAAGAAGVGVPFGILLISYSTAAGVGEG--KLIPLNKNLKSVON 220
DBB     :::: |           :|       :|       :|       :|       :|       :|       :|
        317 KYENGLEKENRELLGS LK-----EIQEATIQSGNSALS KLNKPNLEN    362
QQ      221 PFTTIL-----SNTVKOANKIDDAKLKTTEIAAIGIKETE-----TRRFVV 264
DBB     :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|
        363 IHNCANLRKSEAHSWSQXKWVEINDYLQLQSKEAALKVELELCRSSTAKMRL 422
QQ      265 DYDDLMLSLL-----KEAAKMINTCNEY----QRKHG 293
DBB     :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|   :|
        423 QYEETIMFLVLRTVSQAQRS LANAKDKQIKDEKREG 460

RESULT 8
S18199 myosin heavy chain - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 02-Feb-2001
C/Accession: S18199 Camoretti-Mercado, B.; Perlman, D.; Gupta, M.; Jakovcic, S.; Zak, R.
R;Stewart, A.F.R.; Camoretti-Mercado, B. J. Mol. Evol. 33, 357-366, 1991
A>Title: Structural and phylogenetic analysis of the chicken ventricular myosin heavy ch
A:Reference number: S18199; MUOID:92130260; PMID:1774788
A/Accession: S18199
A>Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1039 <STE>
A/Cross-references: EMBL:X59552; NID:g62995; PIDN:CMA42130.1; PID:g62996
A>Note: In the authors' translation 45-Lys is shown after residue 40, and, consequently,
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; nucleotide binding motif A (p-loop)

```

Query Match	8.0%;	Score 120;	DB 2;	Length 1039;
Best Local Similarity	19.0%;	Pred. No. 7.8;		
Matches	62;	Conservative 73;	Mismatches 132;	Indels 60; Gaps 10;
Qy	1	MTEIVADKTVVVKNAJETADGALDLYNK--YLDQVWPWT-PDETIKELSRPQEQYSOA	57	
Db	369	MTRLMDLLTTQTKLQSGENGFEVQLQEKESLISQLSRGKTSFTQQIEELRLQLEBEETKS	428	
Qy	58	ASVLVGDIKT-----LIMDSQDKFYFATQTVYVCGVATQLLAAYILLFD-----EYNEK	107	
Db	429	KNALAHALQAARHDCDLLREQVEEQAELQRLASKGNAEVAQWRTKYETDAIQRTTEE	488	
Qy	108	KASAKQOILIKV-----LDGCIYKLNBAQKSLIVSSSPNNAASKULLAL	151	
Db	489	LEDAAKKLLARLQBAEEAIEAANAKCSSLEAKHRLQNEQEDMMIDLEKANSAA---SL	545	
Qy	152	D-----SOLFNDRSEKSSYFQSQVDKTRKAYAGAAAGVAGVAGPGLIISYSIAAGVVEG	205	
Db	546	DKQRGFDKIIDNMKQKTEESQAELEASQKE-----ARSLSTELFKKNAYEETLDHLE-	599	
Qy	206	KLIPELKNKLSVQNFFTTLSNTVKAQNKDI-----DAAKULKLTETIAAIGEI	253	
Db	600	-----TLKEENKMLQEEISDLTNQISEGNKNLHEIEKVKKQVEQEKSEVQLALEEAEGALE	655	
Qy	254	KTETETTFYVDYDDMLSLKAEAAKK	280	
Db	656	HEESKTLRFQLELSQKADFERKLAEK	682	

RESULT 9
AH1369
transmembrane protein [imported] - *Listeria monocytogenes* (strain EGD-e)
C:Species: *Listeria monocytogenes*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH1369

S61477
myosin II heavy chain, non-muscle - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C:Accession: S61477; S65349
R:Mansfield, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
J. Mol. Biol. 255, 98-109, 1996
A:Title: Molecular organization and alternative splicing in zipper, the gene that encodes myosin II heavy chain [imported] - Arabidopsis thaliana
A:Reference number: S61477; MUID:96144835; PMID:8568878
A:Accession: S61477
A:Molecule type: DNA
A:Residues: 1-2057 <MAN>
A:Cross-references: UNIPROT:Q94987; EMBL:U35816
R:Mansfield, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
submitted to the EMBL Data Library, September 1995
A:Reference number: S65349
A:Accession: S65349
A:Molecule type: DNA
A:Residues: 1-1908, 'NL', 1911-2057 <MAN>
A:Cross-references: EMBL:U35816; NID:g1141789; PIDN:AAB09049.1; PID:g1572481
C:Genetics: zip
A:Cross-references: FlyBase:FBgn0005634
A:Introns: 10/3; 90/3; 162/3; 215/1; 255/3; 263/3; 303/3; 334/3; 1248/1; 1349/3; 1536/3;
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: alternative splicing; ATP; nucleotide binding; P-loop
F:135-855/Domain: myosin motor domain homology <MMOT>
F:225-232/Region: nucleotide-binding motif A (P-loop)

Query Match 7.6%; Score 115; DB 2; Length 2057;
Best Local Similarity 21.3%; Pred. No. 38;
Matches 80; Conservative 48; Mismatches 107; Indels 140; Gaps 16;

Qy 14 KNAIETADGALDLYNKL-----DQVTPWTFDTIKELSRFK 51
Db 1303 KTVLEKAKGTLEAENADLATELRVNSRQENRRRKAESIAELQV---KLAIEIRAR 1359
Qy 52 QEYSQAASVLVDIKTLMSQDKYFEATQTVVWCGVATQLLAAAYILLFDEYNEK----- 107
Db 1360 SELQEKCTKLOQEAENITNQLAEALKASAAVKSANMSQLTEAQQLLLEEETRQKLGLS 1419
Qy 108 -----KASAAQKDI-LIKVLDDGIT 125
Db 1420 SKLRQISEKEALQOELEDDAKRNVYERKLAETVTTQWQETKKAEDADLAKELSGKK 1479
Qy 126 KLN-----EAQ-KSLIVSSQSFNNSGKL-----LALDSQLTNDFS-EKSSYFQSQ 169
Db 1480 RLNKDIEALERQVKELIAQNDRLDKSKKIQSELEADATIEAQRKTVLELEK---QKN 1536
Qy 170 VDKIRKAYAGAAVGVAGPGLIISYSIA-----AGVVEGKL----- 207
Db 1537 FDKILAEKA-----ISQIAQERDTAREAREKETKVLVSRLDEAFDK 1582
Qy 208 IPELKNLKSQVNFPTLSNTVKQANK---DIDAAKLKLTTEIAAIGETKTETTRFYV 264
Db 1583 IEDLENKRTKLQNELDLANTQGTADKNVHELEKAKALSSQLA---ELKAQNEELE--- 1636
Qy 265 DYDDLMLSLKKEAAK 279
Db 1637 --DDLQLT---EDAK 1646

RESULT 17
F84730
probable myosin heavy chain [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84730
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84730
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1269 <STO>
A:Cross-references: GB:AE002093; NID:g6598483; PIDN:AAC69932.2; GSPDB:GN00139
C:Genetics:
A:Gene: At2g32240
A:Map position: 2

Query Match 7.6%; Score 114.5; DB 2; Length 1269;
Best Local Similarity 18.0%; Pred. No. 22;
Matches 56; Conservative 63; Mismatches 123; Indels 69; Gaps 9;

Qy 1 MTEIVAD--KTVVVKNVIAETADGALDLYNKLQVTPWTFDTIKELSRFKQYEQAA 58
Db 825 LTKLRDLGKIKSYEBQLAEASGSSSLKEKLEQTLGRLLAAASVNE--KLQEFQDAQ 882
Qy 59 SVLVGDIKTLMSQDKYFEATQTVVWCGVATQLLAAAYILLFDEYNEKKAQKDIILK 118
Db 883 E-----KSQSSSESELLAETNNQLK---IKIQLEGLI-----GSGSVKEETALK 925
Qy 119 VLDDGITKNEAOKSLIVSSQSFNNSGKLLALDSQLTNDFSSEKSYFQSQVDKIRKEAY 178
Db 926 RLEEAIERFNQKETE-----SSDLVEKLKTHENQIEBYKKLAH 963
Qy 179 AGAAAGVAGPGLIISYSIAAGVVEGKLPELKNLKSQVNFPTLSNT-----VKQA 232
Db 964 E-----ASGVADTRKY-ELEDALSKLNLESTIEELGAKCQGLEKE 1003
Qy 233 NKDIDAAKLKLTTEIAAIGETKTETTRFYVDVDDMLSLKKEAAKMTNCEYQKRH 292
Db 1004 SGDLAEVNLKLNLELHNGSEANELQTKLSALEAKEQTANELEASKTTIEDLTQLTSE 1063
Qy 293 GKTKLFEVPEV 303
Db 1064 GEKLSQIEKL 1074

RESULT 18
B84800
probable alpha-carboxyltransferase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: B84800
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84800
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-796 <STO>
A:Cross-references: UNIPROT:Q7XJRI; GB:AE002093; NID:g4895181; PIDN:AAD32768.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g38040
A:Map position: 2

Query Match 7.6%; Score 114; DB 2; Length 796;
Best Local Similarity 24.9%; Pred. No. 13;
Matches 68; Conservative 44; Mismatches 119; Indels 42; Gaps 10;

Qy 42 ETIKELSRFKQYEQAAASVLVGIKTLMSQDKYFEATQTVVWCGVATQLLAAAYILLF 101
Db 471 EQILKAKETSTEAEPSPSEVLNEMIEKLKSIDDEYTEATAV---GLESLTA---MR 522
Qy 102 DRYNEKKAQKDIILKVLDDGITKNEAOKSLIVSSQSFNNSGKLLAL-----DS 153
Db 523 EEPF--KASSEHLMHPVLTKIEKLEKEFNTRLTDPAPNYESLKSLLNMLRDRFRAKAS 580
Qy 154 QLTFNDFSSEKSYFQSQVDK-----IRKEAYAGAAAGVAGPGLIISYSIAAGVVE 204

Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEE>
A;Cross-references: GB:M34281

Query Match 7.6%; Score 114; DB 2; Length 2401;
Best Local Similarity 19.4%; Pred. No. 54;
Matches 80; Conservative 57; Mismatches 141; Indels 134; Gaps 16;

QY 5 VADKTVE---VYKNAETADGALDLYNKYLDQVTPWQTFDET--IKELSRKQEV---S 55
Db 1760 IEDKIIKNGLINKLIETKDCMLFYTKTLVETLIKTKTDTYKTSATKSKFLKYID 1819
QY 56 QAAVILVGDITKLLMDSOOKY-----FEATQTVYEWCG 88
Db 1820 ATSNLNDIDNTL-----QTKYDLNQLNHVASMADATNDNNNLLIEKEKATKINN--- 1872
QY 89 VATQLLA-----AYILLFDEYNEKKAQKIDILIKVLDGDTKLNEAOKS 133
Db 1873 -LTELFTDSNKIDAGLHNNKIQIIFENSELHKISIDSIKOLYKQKHAFLKLNIGHNKK 1931
QY 134 LLVSSQSFNNAGKLLALDLSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVAGPGL- 192
Db 1932 YFDISKEDFNI---LQLOESELNLDLKEIGQKISDK--KKQFLHALSETPIFNFTL 1986
QY 193 -----IISYSTAAGVVEG-----KLPELKNKLKSVONFPTTL---SNTVK 230
Db 1987 KEIYHDIVKYKQIDEIENITNEENENITLYMDIITKLKMKVQSILNFVTVYENDSNIIK 2046
QY 231 ---QANKDIDAAKLKLTTT-----IAALGEIKT----- 255
Db 2047 QHIQDNNNDVSKIKESLETTIQSFQKILNKLNEIKAAQFYDNNNNINNVISTISQDVNVK 2106
QY 256 -----ETETRFYVDYDDLMLLS---LLKEAAKMTINTCNEYKRGKK 295
Db 2107 KHISKDLTIENELIQKSLIEDIKKSTYDIRSEQITKYVNPIDHYVEQOTKK 2158

RESULT 21
S10056
hemolysin A - Escherichia coli plasmid pHLy152
C;Species: Escherichia coli
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 04-Mar-2000
C;Accession: S10056
R;Hess, J.; Wels, W.; Vogel, M.; Goebel, W.
FEMS Microbiol. Lett. 34, 1-11, 1986
A;Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparis
A;Reference number: S07209
A;Accession: S10056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1024 <HES>
A;Cross-references: EMBL:M14107
C;Genetics:
A;Genome: plasmid pHLy152
C;Superfamily: hemolysin A; hemolysin A homology
C;Keywords: lipoprotein
P;247-792/Domain: hemolysin A homology <HLYA>
P;564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.5%; Score 113.5; DB 2; Length 1024;
Best Local Similarity 21.4%; Pred. No. 20;
Matches 70; Conservative 59; Mismatches 121; Indels 77; Gaps 14;

QY 2 TEIVADKTVEYVK-----NAJETADGALDLYNKYLDQVTPWQTFDEIKELSRKQEV 53
Db 155 TALSSMKIDELIKQKSGNVSSSELAKASIELINQLVDIVASLN-----NNVNSFSQQ 208
QY 54 YQAAASVL-----VGDITKLLMDSQDKYFEATQTVYEWCGVATQLLAAVILL-FDEY 104

Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEE>
A;Cross-references: GB:M34281

Query Match 7.6%; Score 114; DB 2; Length 2401;
Best Local Similarity 19.4%; Pred. No. 54;
Matches 80; Conservative 57; Mismatches 141; Indels 134; Gaps 16;

QY 5 VADKTVE---VYKNAETADGALDLYNKYLDQVTPWQTFDET--IKELSRKQEV---S 55
Db 1760 IEDKIIKNGLINKLIETKDCMLFYTKTLVETLIKTKTDTYKTSATKSKFLKYID 1819
QY 56 QAAVILVGDITKLLMDSOOKY-----FEATQTVYEWCG 88
Db 1820 ATSNLNDIDNTL-----QTKYDLNQLNHVASMADATNDNNNLLIEKEKATKINN--- 1872
QY 89 VATQLLA-----AYILLFDEYNEKKAQKIDILIKVLDGDTKLNEAOKS 133
Db 1873 -LTELFTDSNKIDAGLHNNKIQIIFENSELHKISIDSIKOLYKQKHAFLKLNIGHNKK 1931
QY 134 LLVSSQSFNNAGKLLALDLSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVAGPGL- 192
Db 1932 YFDISKEDFNI---LQLOESELNLDLKEIGQKISDK--KKQFLHALSETPIFNFTL 1986
QY 193 -----IISYSTAAGVVEG-----KLPELKNKLKSVONFPTTL---SNTVK 230
Db 1987 KEIYHDIVKYKQIDEIENITNEENENITLYMDIITKLKMKVQSILNFVTVYENDSNIIK 2046
QY 231 ---QANKDIDAAKLKLTTT-----IAALGEIKT----- 255
Db 2047 QHIQDNNNDVSKIKESLETTIQSFQKILNKLNEIKAAQFYDNNNNINNVISTISQDVNVK 2106
QY 256 -----ETETRFYVDYDDLMLLS---LLKEAAKMTINTCNEYKRGKK 295
Db 2107 KHISKDLTIENELIQKSLIEDIKKSTYDIRSEQITKYVNPIDHYVEQOTKK 2158

RESULT 21
S10056
hemolysin A - Escherichia coli plasmid pHLy152
C;Species: Escherichia coli
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 04-Mar-2000
C;Accession: S10056
R;Hess, J.; Wels, W.; Vogel, M.; Goebel, W.
FEMS Microbiol. Lett. 34, 1-11, 1986
A;Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparis
A;Reference number: S07209
A;Accession: S10056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1024 <HES>
A;Cross-references: EMBL:M14107
C;Genetics:
A;Genome: plasmid pHLy152
C;Superfamily: hemolysin A; hemolysin A homology
C;Keywords: lipoprotein
P;247-792/Domain: hemolysin A homology <HLYA>
P;564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.5%; Score 113.5; DB 2; Length 1024;
Best Local Similarity 21.4%; Pred. No. 20;
Matches 70; Conservative 59; Mismatches 121; Indels 77; Gaps 14;

QY 2 TEIVADKTVEYVK-----NAJETADGALDLYNKYLDQVTPWQTFDEIKELSRKQEV 53
Db 155 TALSSMKIDELIKQKSGNVSSSELAKASIELINQLVDIVASLN-----NNVNSFSQQ 208
QY 54 YQAAASVL-----VGDITKLLMDSQDKYFEATQTVYEWCGVATQLLAAVILL-FDEY 104

Db 209 LNTGSLNNTKHLNGVN-KLQPLNLDNIGAGLDTV- - - - -SGLSAISAFILSNADAD 264
Qy 105 NEKASAKOILIKVLDD- - -GITKLEAKQSLVSSQSFNAGSKLLALDLSQTLNDFSE 161
Db 265 TRTKAAGVELTTKVLNGVKGISQYIAQR- - - - -AAQGLSTSAAGLAIASAVTLAISP 320
Qy 162 KSSYFQSOVDKIRK- - - - -EYAGAAAGVAGVGPGLIISYSIAAGVVGKLIPELKNKLKS 217
Db 321 LS- - -FLSIADKPKRANKIEYSORPKLGVGDGDSLLAAPHKETGAIDASL- - - - - 368
Qy 218 VQNFPTLSNTVKQANKIDDAKLLKLTETI- - - - -AAIGEIKTETETTRFYVDYD 267
Db 369 - - - - -TTISTVLASVSGISAA- - - - -ATSLVGAPVSLVGNVGIISGLEASK- - - - - 414
Qy 268 DLMLSLKEAKAKMINTCNEYQKRHK 294
Db 415 - - -QAMFEHVASKWADVIAEWKKGK 438

RESULT 22
T22716
hypothetical protein F55C5.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T22716
R;Harris, B.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19604
A;Accession: T22716
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-622 <WIL>
A;Cross-references: UNIPROT:Q20822; EMBL:Z78198; PIDN:Z78198; GSPDB:GN00023; CESP:F55C5
A;Experimental source: clone F55C5
C;Genetics:
A;Gene: CESP:F55C5.8
A;Map position: 5
A;Introns: 24/3; 44/2; 147/3; 251/3; 561/3

Query Match 7.5%; Score 113; DB 2; Length 622;
Best Local Similarity 20.6%; Pred. No. 11;
Matches 67; Conservative 54; Mismatches 100; Indels 104; Gaps 15;

Qy 41 DETIKLSRFKQYSQAASVLGVGDIKTLMDSDQKFEATQTVYEW- - - - -CGVAT 91
Db 251 DKLISEM-RASATSAEVVTIEWGAKSTVDBE- - - - -KAKQVQEWKQTEVELAQCCQTPK 304
Qy 92 QLLAAAYILLDFEYNEKASAKQDILIKVLDDGIGTKLNE-AOKSLIVSSQSF- - - - -NNA 144
Db 305 EKWALP- - - - -EKATADTRDAIDRISDIIRKSSSENADTTVLQSIKAYLEFLKMGNT 356
Qy 145 SGKLLALDSQTLNDFSEKSSYFQS- - - - -QVDKIRKE 176
Db 357 ASRYLAI- - -IDNTKSEKSKPQDLRLYDSVIEIYKEVABIPGADHKNLIQAFVKVE 413
Qy 177 AYAGAAAGVAGVGPGLIISYSIAAGV- - - - -VEGKLIPELKNKLKS- - - - -VQ 219
Db 414 YYRAFRCFYMASSYSALHKYSEAAALPDRTVRSRVQDAEGKL- - - - -KKLKSSSFITNETQ 468
Qy 220 NFFTLLSNTVKQANKIDDAKLLKLTETIAGETITETTRFYVDYDMLMLLKRAAK 279
Db 469 SSLNELRSEVESAKVTVRAARL- - - - -ASAAGDKTDSSELAKEI-ID- - - - -XR 510
Qy 280 KMINTCNEYQK- - -RHGKTLFEVP 301
Db 511 PLLETVNEWRQWVRNLSLKOKKTIIP 535

RESULT 23
A35254
leukotoxin A - Pasteurella haemolytica (serotype T10)
N;Alternate names: lktA protein

C;Species: Pasteurella haemolytica
C;Date: 10-Aug-1990 #sequence_revision 15-Nov-1996 #text_change 09-Jul-2004
C;Accession: S37145; A35254; S34237; S34235
R;Lainson, A.F.; Aitchison, K.; Donachie, W.
submitted to the EMBL Data Library, September 1993
A;Description: DNA sequence of the leukotoxin A gene from P. haemolytica T10 serotype.
A;Reference number: S37145
A;Accession: S37145
A;Molecule type: DNA
A;Residues: 1-955 <LAI>
A;Cross-references: UNIPROT:P55117; EMBL:Z26247; PIDN:CAA81206.1; PID:G4004
R;Highlander, S.K.; Engler, M.J.; Weinstock, G.M.
J. Bacteriol. 172, 2343-2350, 1990
A;Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.
A;Reference number: A35254; MUID:90236888; PMID:2185213
A;Accession: A35254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 950-955 <HIG>
A;Cross-references: GB:M24197; GB:M34943; GB:M34944
R;Lainson, A.F.; Aitchison, K.D.; Donachie, W.
submitted to the EMBL Data Library, June 1993
A;Description: DNA sequence of the carboxy terminal end of leukotoxin A from the T3 sero
A;Reference number: S34235
A;Accession: S34237
A;Molecule type: DNA
A;Residues: 745-955 <LA2>
A;Cross-references: EMBL:Z22884; NID:g311828; PIDN:CAA80498.1; PID:g311829
A;Experimental source: serotype T3
A;Accession: S34235
A;Molecule type: DNA
A;Residues: 723-955 <LA3>
A;Cross-references: EMBL:Z22887; NID:g311824; PIDN:CAA80501.1; PID:g311825
A;Experimental source: serotype T10
C;Function:
A;Description: attacks cell membranes and causes cell lysis
C;Superfamily: hemolysin A; hemolysin A homology
C;Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;
F;240-786/Domain: hemolysin A homology <HLYA>
F;718-809/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIIYF]-X)
F;718-726/Region: repeat
F;727-735/Region: repeat
F;736-744/Region: repeat
F;745-753/Region: repeat
F;754-762/Region: repeat
F;763-771/Region: repeat
F;772-780/Region: repeat
F;781-789/Region: repeat
F;792-800/Region: repeat
F;801-809/Region: repeat
F;556/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.5%; Score 113; DB 1; Length 955;
Best Local Similarity 18.6%; Pred. No. 19;
Matches 69; Conservative 67; Mismatches 136; Indels 98; Gaps 12;

Qy 6 ADKTEVEVKNATADGALDLYNKYLPQVTPWOTFDETIKLSRFKQYSQAASVLVG- - 63
Db 89 AQTSLGTIQNVLGITERGIVLSAPQLDKLQKNKVGQALGSSESTAQNFSQAKTVLSGVQ 148
Qy 64 - - - - -DIKTLMDSDQKY- - - - -FEATQTVYEWGVATQLLAAVILLDFEYNEK 107
Db 149 GNSRTVLAGMDLDEALQNESDQLTLAKAGLELTNSLTIANSVQTLDAFSEQISQFSGK 208
Qy 108 KASAQKDILIKVLDD- - -GITKLEAKQSLIVSSQSFNAGSKLLALD- - - - - 152
Db 209 LQNVKG- - -LGAJGDKLKNITGGLDKAGLGDVVKRLSGATAALVLADKDASTAKKVGAG 265
Qy 153 - - - - -SQTNDFFSEK-SSYFQSOVDKIRKEAYAGAAAGVAGV- - -GPFGLIISYSIAAGV- - -VE 204
Db 266 PELANQVGVNITRAVSSVILAQ- - - - -RVAAGLSSTGTPVAALIASVVAISPLS 315
Qy 205 GKLIPELKNKLKSVQNPF- - - - -TTLSNTVKQANKIDDAKLLKLTETIAAI- - - - - 250

```
Db 316 FAGIADKFRKSLNTAERFKKLGEGDSLLAEYQHGCTTIDASVTAINALAAAGGV 375
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Qy 251 -----GRIKTETTRFYVDYDDMLSLKBAKAKMINTCNEYQ 289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 SAAAGSVASPIALLVSGITGVISTILOYSK-----QAMFEHVANKIHNKIVEWE 426
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 290 KRHGKKTLE 299
      | : : : |||
Db 427 KNNGGKNYFE 436
      | : : : |||

RESULT 24
S48385
hypothetical protein YII149c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C:Accession: S48385
R:Churcher, C.
submitted to the EMBL Data Library, September 1994
A:Reference number: S48310
A:Accession: S48385
A:Molecule type: DNA
A:Residues: 1-1679 <CHU>
A:Cross-references: UNIPROT:P40457; GB:Z47047; EMBL:Z38059; NID:g603997; PID:g763197; CS
C:Genetics:
A:Gene: SGD:MLP2; MIPS:YII149C
A:Cross-references: SGD:S0001411
A:Map position: 9L

Query Match 7.5%; Score 112.5; DB 2; Length 1679;
Best Local Similarity 20.6%; Pred. No. 42;
Matches 51; Conservative 44; Mismatches 67; Indels 85; Gaps 10;

Qy 39 TFDTEIKELSR-----FKQEYSQAASVLVDIKTLMSQDKYFEATQVTEWCQVATQLL 94
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1134 SLEKTINDLQRTQLTSEKEY-QCSAVIIDFKDITK-----EVTQV----- 1173
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 95 AAYILLFDEYNEKKAQKDI-----LIKVLDDGITKLNEAOKSLVSSQSFNNAS 145
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1174 -----NILKENNAILOKSLNVTENKREIYKQNDQOEISRLQRLDIOTKQSVINS 1226
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 146 GKLLALDSQLT-----NDFSEKSYFSQ-QVDKIRKEAYAGAAAGVAGVPGFLIISYI 198
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1227 NKILVSEMEQCKORYQDLSSQKQDAQKDKIEKLTNE----- 1264
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 199 AAGVVEGKLIPELKNLKSQVNFITLNTV-----KQAKKDIDAALKL---LTTEIAAIG 251
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1265 -----ISDLKGLSSAENANADLENKFNRLKQAEKLDASKKQQAALTNELNELK 1315
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 252 EIKTETE 258
      | : : : |||
Db 1316 AIKDKLE 1322
      | : : : |||

RESULT 25
G97827
hypothetical protein RC1023 [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: G97827
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: G97827
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <KUR>
A:Cross-references: UNIPROT:Q92GU9; GB:AE006914; PIDN:AA03561.1; PID:g15620140; GSPDB:G
C:Genetics:
A:Gene: RC1023
```

```
Query Match 7.4%; Score 112; DB 2; Length 295;
Best Local Similarity 24.9%; Pred. No. 5;
Matches 56; Conservative 40; Mismatches 83; Indels 46; Gaps 11;

Qy 94 LAAYILLFDEYNEKKAQKDIILIKVLDDGITKNE--AQKSLVSSQSFNNASGKLAL 151
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 4 LLLIITVFTFVNAQSL--PNIVASVNDPEIT-LNEFRARKKMTM---ALNNVESLTPAQ 58
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 152 DSQLTN-----DFSEKSYFSQVDK-IRKEAYAGAAAGVV---AGPFGILLIISYIAAGVV 203
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 59 DKQLSDIALKSLIDESLLFOYAGDREIQQEETENAKSIEDRNKMPHGSLQLYKLSRVN 118
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 204 EGKLIPELKNLKSQVNFITLNTVQKQAKDIDAALKLTTETIAAIGIKETETTRY 263
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 119 PDSFTSQIKSELIKM-NILSSLSRSVQVSNKEIDVAILL----- 156
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

Qy 264 VYDDLMLSL-----LKEAAKKNMINTCNEYQKR-----HGKKTLE 299
      | : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 157 SPQKQVSEISMVFTSKDGGNKAPTQNNLNKRLKCKADVKKSLYD 201
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 26
JC4298
hyaluronan receptor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
C:Accession: JC4298; A42925; A41923; S21586
R:Entwistle, J.; Zhang, S.; Yang, B.; Wong, C.; Li, Q.; Hall, C.L.; A, J.; Mowat, M.; Gr
Gene 163, 233-238, 1995
A:Title: Characterization of the murine gene encoding the hyaluronan receptor RHAMM.
A:Reference number: JC4298; MUID:96011639; PMID:7590272
A:Accession: JC4298
A:Molecule type: mRNA
A:Residues: 1-631 <ENT>
A:Cross-references: UNIPROT:Q00547; EMBL:X64550
A:Experimental source: 3T3 fibroblast
R:Hardwick, C.
J. Cell Biol. 118, 753, 1992
A:Reference number: A42925; MUID:92348516; PMID:1639856
A:Contents: erratum
A:Accession: A42925
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 155-376, 'S', 378-504, 'E', 506-631 <HAR>
A:Cross-references: GB:X64550
A:Note: authors translated the codon AGT for residue 377 as Thr and CTA for residue 507
R:Hardwick, C.; Hoare, K.; Owens, R.; Hohn, H.P.; Hook, M.; Moore, D.; Cripps, V.; Auste
J. Cell Biol. 117, 1343-1350, 1992
A:Title: Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motil
A:Reference number: A41923; MUID:92299690; PMID:1376732
A:Accession: A41923
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 155-376, 'S', 378-504, 'E', 506-507, 'L', 508-630 <HA2>
A:Cross-references: GB:X64550
A:Note: this sequence has been corrected in reference A42925
C:Comment: This protein regulates cell motility and transformation, and focal adhesion d
C:Genetics:
A:Gene: rhamm
A:Introns: 20/3; 54/2; 79/2; 137/3; 187/3; 259/2; 382/2; 431/2; 482/2; 515/3; 574/3; 625/
C:Superfamily: hyaluronan receptor
C:Keywords: glycoprotein; receptor
F:260-382/Region: 21 residue repeats
F:516-574/Region: hyaluronan binding #status predicted
F:575-625/Region: hyaluronan binding #status predicted
F:116,283,304,325,346,367,398,438,619/Binding site: carbohydrate (asn) (covalent) #statu

Query Match 7.4%; Score 112; DB 2; Length 631;
Best Local Similarity 20.5%; Pred. No. 13;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

Qy 1 MTEIVADKTVEVKNNAIETADGALDLYNKYLDQVPMQTFDETIKELSRFKQEYSQAASV 60
      : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
```

Db 253 LDNLLREKEVELEKHAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309
Qy 61 LVGDITKLLMDSQDKYFEATQTVYVWCGVATQALAAAYILLDFDYNEKASQAQKILIKVL 120
Db 310 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQSLRDVTAQ-L 359
Qy 121 DDGITKLENAQKSL-LVSSQSFNNASGKLAL-DSQLTN-DFSEKSYFQSQVDKIRKEA 177
Db 360 ESVOEKYNDTAQSLRDVTAQLESYKSTLKEIBDLKLENLTLOEKVAMAEKSVEDVQQOI 419
Qy 178 YAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKKN-----LKSQVQFF---TTL 226
Db 420 LTAESTNQ-----EYA-----RVQDLQNRSTLKEEIKETSPLEKITDLK 462
Qy 227 NTVKQANKOI-----DAKLKLTTRIAIGETTTETTRFRFYVDYDL 269
Db 463 NQLRQODEPRKOLEKGRKTAENVMTELT-----MEINKWRLLYDEL 507
RESULT 27
S38173
A;Molecule type: DNA
A;Residues: 1-1875 <BAL>
A;Cross-references: UNIPROT:Q02455; EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR095w
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jil
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of *Saccharomyces cerevisiae* chromo
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S31207
A;Molecule type: DNA
A;Residues: 1-300, 'A', 302-1875 <KOE>
A;Cross-references: EMBL:L01992; NID:gl71958; PIDN:AAA34783.1; PID:gl71959
A;Gene: SGD:MLP1
A;Cross-references: SGD:S0001803; MIPS:YKR095w
A;Map position: 11R
Query Match 7.4%; Score 112; DB 2; Length 1875;
Best Local Similarity 19.7%; Pred. No. 52;
Matches 68; Conservative 66; Mismatches 122; Indels 90; Gaps 15;
Qy 2 TEIVADKTVVKNNAIETAGALDLYNK---YLDQVI-----PWQTFDETIK----- 45
Db 685 SRLAERFKLNTLTDLTKAENDQLKRPDYLQNTILKQDSKTHETINYSVCSKLSI 744
Qy 46 ---ELSRFKQEV-----SQAAVLVDGITKLLMDSQDKYFEATQ 82
Db 745 VETELLNKEQKLVHLEKNLQKLNKLSPEKDSLRIMWTQTLQTKEREDLLEETRKS 804
Qy 83 VYWCGVATQALAAAYILLDFDYNE-----KKAQAQKILIKVL-DDGITKLENAQKSLV 136
Db 805 CQK-----KIDELEDALEKSKTSQKDHIIKQLEEDNNSNIWQNKIEA 850
Qy 137 SSQSFNWASGKLALDSQLTN-----DFSEKSYFQSQVDKIRKEAYAGAAAGVAGPFG 192

Db 851 LKDYISV---ITSVDSKQTDIEKLYKVKLSKEIEEDKIRLHTY-----NVM-----ET 899
Qy 193 IISYSIAAGVVEGKL-IPELKNKLSQVNFPTTSLNTVKQANKDIDAQKLTTEIAAIG 251
Db 900 INDSLRLKELEKSKINLTDAYSQIKEYKDIYETTSQSLQQTNSKLDSEKDFDTNQIKNLT 959
Qy 252 EIKTETTRFRFYVDYDLMLSLLEKAAKQKMTICNE--YKRRHGKK 295
Db 960 DEKTSLED-----KISLKE---QMFLNLLNLDLQKGMKEK 992
RESULT 28
S67087
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-TKNLQKQMDLANNLNDNL 536
Qy 175 KEAYAGAAAGVAGPFGELIISYSGAAGVVEGKLIPELKN-----KLKSVQNFPTTSLNTVK 230
Db 537 K-----VQDKLI---KNEETLKLKEA-----ID 558
Qy 231 QANKDIDAQKILTT--ETAAIGETITET-----ETTRFYVDYDL----- 269
Db 559 SLNSEMDLKKQITSKDDDEFKMQSKYETVEDEAKTRNAEVTENGDIEDLKSKSLHEE 618
Qy 270 MLSLLEKAAKQKMTICNEVQRRHGKTLFEV 300
Db 619 TITELNKHVHLENEC-ELEKQKFEKTSLEL 648
RESULT 29
T18296
A;Molecule type: DNA
A;Residues: 1-821 <HUG>
A;Cross-references: UNIPROT:Q08581; EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:SLK19; MIPS:YOR195w
A;Cross-references: SGD:S0005721
A;Map position: 15R
Query Match 7.4%; Score 111.5; DB 2; Length 821;
Best Local Similarity 21.8%; Pred. No. 20;
Matches 72; Conservative 52; Mismatches 112; Indels 95; Gaps 14;
Qy 10 VEVVNAIETAGALDLYNKLYLDQVFPWQTFDETIKLSRPFQKQYSAASVL----- 61
Db 373 VEKFKRIKRELNTEIKVLNS--NOKILOEKFDASITEVNHKIGEHENTVNTLQONEKILN 430
Qy 62 -----VGDITKLLMDSQDKYFEATQTVYVWCGVATOLAAAYILLDFDYNEKASQAQKIL 116
Db 431 DKVVELNEMKAEIKGNNDKLSYEYETINDLNSRIVQL-----NDKIEST--DIV 477
Qy 117 IKVLDPGITKLENAQKSLVSSQSFNNAS--GKLLALDSQLTNDFSEKSYFQSQVDKIR 174
Db 478 LKSKENELDNKLSLKTETLSISKDFNDSDLIGQINELIS-T

A;Residues: 1-2139 <GUI>
A;Cross-references: UNIPROT:Q07569; EMBL:L03534; MID:g1850912; PID:g1850913; PIDN:AAB480
C;Genetics:
A;Gene: mhca
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 7.4%; Score 111.5; DB 2; Length 2139;
Best Local Similarity 21.3%; Pred. No. 67;
Matches 74; Conservative 57; Mismatches 128; Indels 91; Gaps 15;

QY 7 DKTEVVKVNAIETAD-----GALDLYNKYLQVPIWQ-----TFDETIELSRFKQY 54
DB 986 DKTTAEQESIDEKEDITKLGDKILBEEKDLEQDRADVSNATKDDIAKLNKTIETC 1045
QY 55 SQAASVLVGDITLLMDSQKIFPATQTVIEW-----CGVATQLLAAYILLFDEYNEKASA 111
DB 1046 EDARDE-IAKLELEDEENKNDLTNELQOTQLKGETEKSAAQVAA-----TKASD 1099
QY 112 QKDILIKVLD-----GITKLN-EAOKSLVSSQSF-----NNASGKILA 150
DB 1100 ERDLSQNLNEKUTTNLTITKADLEKKISGLKQDYEDLEDKDKKIEGLRQAQRIKE 1159
QY 151 LDSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVAGPFGLLIISYSIAAGVVEGKLI-P 209
DB 1160 LDDEITKG-ADVSQYLQKQ-----KEEYES-----QIAKNQBEKEAIGN 1197
QY 210 ELKNKLKS-----VQNFFTLNTVQKANKDIDAAKLKLTITTAIGIKTETETTRF 262
DB 1198 DVKNKEKTIKEKELEIQSLQEKLDDET-EVEKEDAEKKKTEKEKMKALQEBKENVESKN 1256
QY 263 YVDVDDLML-----SLLKEAAKGMINTCNEYQKRHGK 294
DB 1257 STEKDKKLENDKDTQKLDMDTADNEKUKAKAKALEAQLNEVDNHEK 1306

RESULT 30
F87908
Protein T22A3.8 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
A;Accession: F87908
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: F87908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2823 <STO>
A;Cross-references: UNIPROT:Q045614; GB:chr_I; PIDN:CAAL15432.1; PID:g3924779; GSPDB:GN000
A;Accession: F87908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2823 <ST2>
A;Cross-references: GB:chr_I; PIDN:CAB03385.1; PID:g3924881; GSPDB:GN00019; CESP:T22A3.8
C;Genetics:
A;Gene: T22A3.8
A;Map position: 1
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.4%; Score 111.5; DB 2; Length 2823;
Best Local Similarity 21.3%; Pred. No. 95;
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

QY 25 DLVKNKYLDQVPIWQTFDETIELSRFKQYSAASVLVG-----DIKTLMD 72
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADARMSLVVGRINRYKEVSNIEKURVEA 1939
QY 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104
DB 1940 EDQIAYSRSNIEKARSEELMFMFDEKINNTLAELPDLVEQCONITLL-----YSQILDEY 1996
QY 105 NE-----KKASAQKDIILIKVLDGDKITKLEAQAOKSLVSS-----OSFNAS--- 145
DB 1997 DEEYVQTAGRHAELKLEVQAQ-----KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053
QY 146 -----GKLALDSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVAGPFGLLIIS 195
DB 2054 DSAAEASEAVSKMLGSESGSDANEES--LRSQLEKLKNES-----S 2094
QY 196 YSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVQKANKDIDAAKLKLTITTAIGIKT 255
DB 2095 LSNVDNSNAVKIVBELKKEKD-----LTDRGLHLN-----ELK-TSIVKRLGVKN 2140

1940 EDQIAYSRSNIEKARSEELMFMFDEKINNTLAELPDLVEQCONITLL-----YSQILDEY 1996
QY 105 NE-----KKASAQKDIILIKVLDGDKITKLEAQAOKSLVSS-----OSFNAS--- 145
DB 1997 DEEYVQTAGRHAELKLEVQAQ-----KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053
QY 146 -----GKLALDSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVAGPFGLLIIS 195
DB 2054 DSAAEASEAVSKMLGSESGSDANEES--LRSQLEKLKNES-----S 2094
QY 196 YSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVQKANKDIDAAKLKLTITTAIGIKT 255
DB 2095 LSNVDNSNAVKIVBELKKEKD-----LTDRGLHLN-----ELK-TSIVKRLGVKN 2140

256 ETEETRFYVDYDDLMLSLLEAKKGM-INTCNEYOKRHGKKT 296
2141 EASS-----WDDKDHMSILKNGAKTAHARSANVKESEGIKT 2179

RESULT 31
T23064
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T23064; T25096
R;Barlow, K.
submitted to the EMBL Data Library, October 1997
A;Reference number: Z19669
A;Accession: T23064
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2823 <WIL>
A;Cross-references: UNIPROT:Q45614; EMBL:AL008585; PIDN:CAAL15432.1; GSPDB:GN00019; CESP
A;Experimental source: clone H10E24
R;McMurray, A.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19980
A;Accession: T25096
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2823 <W12>
A;Cross-references: EMBL:Z81125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8
A;Experimental source: clone T22A3
C;Genetics:
A;Gene: CESP:T22A3.8
A;Map position: 1
A;Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1989/2; 2760/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.4%; Score 111.5; DB 2; Length 2823;
Best Local Similarity 21.3%; Pred. No. 95;
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

QY 25 DLVKNKYLDQVPIWQTFDETIELSRFKQYSAASVLVG-----DIKTLMD 72
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADARMSLVVGRINRYKEVSNIEKURVEA 1939
QY 73 QDK-----YFEATQTV-----YEWCGVATQLLAAYILLFDEY 104
DB 1940 EDQIAYSRSNIEKARSEELMFMFDEKINNTLAELPDLVEQCONITLL-----YSQILDEY 1996
QY 105 NE-----KKASAQKDIILIKVLDGDKITKLEAQAOKSLVSS-----OSFNAS--- 145
DB 1997 DEEYVQTAGRHAELKLEVQAQ-----KIVDRFVDTRTETENPLKASHAYENIVEALKNATEAV 2053
QY 146 -----GKLALDSQLTNDFSEKSSYFQSDVKIRKEAYAGAAAGVAGPFGLLIIS 195
DB 2054 DSAAEASEAVSKMLGSESGSDANEES--LRSQLEKLKNES-----S 2094
QY 196 YSIAAGVVEGKLIPELKNKLKSVQNFFTLNTVQKANKDIDAAKLKLTITTAIGIKT 255
DB 2095 LSNVDNSNAVKIVBELKKEKD-----LTDRGLHLN-----ELK-TSIVKRLGVKN 2140

QY 256 ETETTRFYVDYDMLSLLEAKKM-INTCNEYQKRGHKKT 296
Db 2141 EASS---WDDKHDRMHSILKNGAKTAHERSANVKKSEGIKT 2179

RESULT 32
T43291
laminin alpha chain - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T43291
R;Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang
submitted to the EMBL Data Library, June 1998
A;Description: Expression, function and evolution of laminin alpha chains.
A;Reference number: 222397
A;Accession: T43291
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3102 <ZHU>
A;Cross-references: UNIPROT:O45614; EMBL:AF074902; PIDN:AAC26793.1
C;Genetics:
A;Map position: 1
A;Note: lamal/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.4%; Score 111.5; DB 2; Length 3102;
Best Local Similarity 21.3%; Pred. No. 1.1e+02;
Matches 73; Conservative 56; Mismatches 100; Indels 113; Gaps 17;

QY 25 DLVNYKLYDVPWQT-FDETIKELSRFKQESQAASVLVG-----DIKTLMDS 72
Db 1881 DLKNR-IDVLEQWMDYRETIYDVSKKDTADAEKMSLVGVGRINRYKEVSNIEKLRVEA 1939

QY 73 QDK-----YFETATQV-----YEWCGVATQLLAAYILLFDEY 104
Db 1940 EDQIAYSRNSTEKARSEELMNFEDKEKNWTLAELPDLVEQCQNITLL---YSQLIDEX 1996

QY 105 NE-----KASAKDILIKVLDDGGITKLENAOKSLVSS-----QSFNAG--- 145
Db 1997 DEEYVOTAGRHAELVQAO---KIVDRFVDTRETENPLKASHAYENIVEALKNATEAV 2053

QY 146 -----GKLLALDSQLTNDPSEKSSYFQSOVDKIRKEAYAGAAGVAGPFGLIIS 195
Db 2054 DSAEASEAVSKMGSGSGSDANEE---LRSQLKLNES-----S 2094

QY 196 YSIAAGVVEGKLIPELKNKLKSVQNFPTTSLNTVKQANKDIDAKLKLTTETIAAIGIKT 255
Db 2095 LSNVDNSNAKIVPELKKKKO-----LTDLRLHNL-----ELK-TSIVKRLGVKN 2140

QY 256 ETETTRFYVDYDMLSLLEAKKM-INTCNEYQKRGHKKT 296
Db 2141 EASS---WDDKHDRMHSILKNGAKTAHERSANVKKSEGIKT 2179

RESULT 33
H88391
protein R06B10.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: H88391
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: H88391
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-934 <STO>
A;Cross-references: UNIPROT:O44874; GB:chr_III; PIDN:AAB95035.1; PID:G2746881; GSPDB:GNC
C;Genetics:
A;Gene: R06B10.2

A;Map position: 3

Query Match 7.4%; Score 111; DB 2; Length 934;
Best Local Similarity 21.6%; Pred. No. 25;
Matches 69; Conservative 43; Mismatches 131; Indels 76; Gaps 14;

QY 5 VADTVEVVKNAIETADGALDLYNKYLDQV-IPWQTDETIKELSRFKQESQAASVLVG 63
Db 81 VYEKTGVPGSIPKPD---DVYKTQDALKIKWSKVDSPAKGIVIFLNEMESMEGVTS 137
QY 64 DIKTLMDSDQKYPEATQTVVYECGATQLLAAYILLFDEYNEKKASAKQDILIKVLDDG 123
Db 138 DDANIKLSDNAPHLPRFK---DW-----TLEEVN----- 163
QY 124 ITKLENAOKSLVSSQSFN---NASGKLL-----ALDSQLTNDPSEKSSYFQSOVDKIRKE 176
Db 164 VYAQNHRLNIHISYPSNLNVYKAVAEFFPKNSLDGSLLPQVQENINKLNTLAEKIKPE 223
QY 177 AYAGAAGVAGPFGLIISYISIAAGVVEGKLIPELKNKLKSVQNFPTTSLNTVKQANKDI 236
Db 224 -----APMISQLHELFLSRSLKKS-VNFKIISGLNLGSIQIMIFDTLKN--EWVNKT 274
QY 237 DAAKKLTTETIAAIGETIKTETTRFYVDYDMLSLLEAKKMKNMINTCNEYQKRGHKKT 296
Db 275 DGQAYNLG---KALNQKLTGNNAQ-QVD-----LSIVPYEEKIFKDVSNLYQKVNLSKD 325

QY 297 L-----FEVPE 302
Db 326 LTVIVRPLIGTVTGIEIPQ 344

RESULT 34

LEECA

hemolysin A - Escherichia coli

C;Species: Escherichia coli

C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 09-Jul-2004

C;Accession: A24433; I41280

R;Feilmee, T.; Pellett, S.; Welch, R.A.

J. Bacteriol. 163, 94-105, 1985

A;Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.

A;Reference number: A24433; MUID:85234404; PMID:3891743

A;Accession: A24433

A;Molecule type: DNA

A;Residues: 1-1023 <PEL>

A;Cross-references: UNIPROT:P09983; GB:M10133; GB:M12863; NID:G146377; PIDN:AAA23975.1;

A;Experimental source: strain J96, O4 serotype

R;Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.

Science 266, 1992-1996, 1994

A;Title: Fatty acylation of two internal lysine residues required for the toxic activity

A;Reference number: A55387; MUID:95099325; PMID:7801126

A;Contents: annotation; lysine palmitoylation

A;Note: lysine modification is performed by the hlyC gene product

R;Haertlein, M.; Schiessl, S.; Wagner, W.; Rdest, U.; Krest, J.; Goebel, W.

J. Cell Biol. 22, 87-97, 1983

A;Title: Transport of hemolysin by Escherichia coli.

A;Reference number: I41280

A;Accession: I41280

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1,'T',3,'V',5,'T',7-44 <RES>

A;Cross-references: GB:M29173; NID:G146337; PIDN:AAA23957.1; PID:G146338

C;Genetics:

A;Gene: hlyA

C;Function:

A;Description: attacks blood cell membranes and causes cell lysis

C;Superfamily: hemolysin A; hemolysin A homology

C;Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;

F;246-791/Domain: hemolysin A homology <HLYA>

F;723-851/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIFY]-X)

P;563,689/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match

Best Local Similarity 21.3%; Pred. No. 28;

7.4%; Score 111; DB 1; Length 1023;

193	Db	DIVLP--AYDSSKLLSFNELQRFERLVAQAATQNNNEEGEKMKFKVPFELSHSDVRTLL	250
70	QY	-----MDSQDKYFEATQTVYEWGV-----ATQLLAAYTLLPDE	103
251	Db	YSLDPTLTKESDSFIADVYDTYF--IVNYSWSDNSDKYFKFNYTRTGTVTSIDFDS	308
104	QY	-----KKASAQKQILIKVLDDGITKLNEAQKSLVSSQSFNNASGKL	148
309	Db	KTEVPMTRNWEVVEPIQSLQKQKDEQIKDLTKQVNIQNKDKVGI---EQQFNTASEKL	365
149	QY	-----LTNDFSEKSSYFQSQVDKIRKEAYNAGAAGVVAGVPGFLI	193
366	Db	VQLNSEVEQLKPYKEKHEKTLLEQKLSERKEFYKARFEALNAEE-----	409
194	QY	ISYSTAAGVWGKLIPELKNKLKSVQNFHTLSNTVKQANKQIDAAKLKLTTETIAAIGBI	253
410	Db	-----KFTSTEVQNL-----IHASVKQ--DEEGEKAVLQNLTMVLVLVS	447
254	QY	KTETETT	260
448	Db	PTETNTT	454

RESULT 36
T44967
gas-vesicle protein gvpC - Halorubrum vacuolatum

C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #date_change 01-Jan-2000
C;Accession: T44967
R;Mayr, A.; Pfeifer, F.
Arch. Microbiol. 169, 24-32, 1997
A;Title: The characterization of the gvpACNORGH gene cluster involved in gas vesicle formation in *Halobacterium salinarum* R1
A;Reference number: Z22885; PMID:97361906; PMID:9211710
A;Accession: T44967
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-540 <MAY>
A;Cross-references: UNIPROT:Q33398; EMBL:Y08571; NID:g2598010; PIDN:CAA69882.1
A;Experimental source: strain DSM 3638
C;Genetics:
C;Gene: gvpC

Query Match	7.3%;	Score 110.5;	DB 2;	Length 540;
Best Local Similarity	20.9%;	Pred. No. 13;		
Matches	71;	Conservative 42;	Mismatches 130;	Indels 97; Gaps 12
QY	5	VADKTVEVV---	KNALETADGALDLYNKVLDQVIEPWQTFDETIKELSRFKOEYSQAASVL	61
DB	3	VNEKGEEMWAKSEKFVEVQGRATTEVAASPADDDVEQKQDTSQVASIAAFREEMQQSGAVF	62	
QY	62	V-----GDIKTL---	LMSDDKQFPEATQTVYEW-----CGVATOLLAA	96
DB	63	ESYSEEPDGDVNRNLNAIEIAEQDDAFATTAFAFEYRESFHQDVQALLTALGEIQQFVE	121	
QY	97	YIILFDYNEKKASAKDILIKVLDGIGIKLNAEAKSLIVSSQSFNNAAGKLLALDLSOL-	151	
DB	123	VEDVFDQVAGDFGEAVDEFNVAEVSQDRIAFENTAAAFEEYSKDFHREKVGALLAAVDLRL	181	
QY	156	-----TNDFSEKSVFSQSDVKIRKEAYAGAAAGVAGPFF-----	191	
DB	183	ARFANVTAAFDQVAEDFEQSVDSEFNSEKDDQD-----AAFEITAAEFSEYRDEPHEQDV	231	
QY	191	-GLIISYSIAAGVVVEGKLIPELANKLKSQNFPTTLLSNTVTKQANKDID--AAKLKLITTE-	241	
DB	217	PGIATA-----IEELSRFSVDVADEGRYAD---BFNGEVDTRASFEDTAA	281	

```

247  -----IAAIGEIKTETTR--FYVDYDD 268
      :||| |:| |:| |:| |:|
282 FEEYDKFHRQEVQAVLAAIDELENDIETATRAEFAATEDD 321

```


bps2 protein homolog (bps2) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: A90394
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: A90394
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-587 <KUR>
A;Cross-references: UNIPROT:Q97WH8; GB:AE006641; NID:gl3815540; PIDN:AAK42408.1; GSPDB:G
C;Genetics:
A;Gene: bps2

Query Match 7.3%; Score 110; DB 2; Length 587;
Best Local Similarity 20.6%; Pred. No. 16;
Matches 59; Conservative 56; Mismatches 106; Indels 66; Gaps 12;

Qy 23 ALDLYNKYLQVTPWQTFDETI--KLSRFKQYSQAASVLVGDITKLLMDSQDKYPEAT 80
Db 51 AEDLLNVFADSGVYEALDNKLYRRKIRNGLGEKKNLMDDRRALL--TFSP 106
Qy 81 QTVYEWCGVATQALAA-----YILLFDEYNEKASQAQKILIKVLDDGTTKLNKAAQKSL 134
Db 107 NR-----LVQTLLSGDGVWFSTTSKINEIK--AKBELQKLTAEINARDELQK-- 156
Qy 135 LVSSQSENN---ASGLLADSQLTNDPSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFG 191
Db 157 -----KYNRIEIOAKIRALD-----BEIDKLEK--RESSNIVAK--- 191
Qy 192 LIISYSIAAGVVEGKIPKLNKLSVQNFTLSTNVKQAKDIDAAKLKTAEIAAIG 251
Db 192 --TYTIT--LTRQNKINEILNKIKVKDELALNFALKIEIEIQNESKVSFDIKTL 247
Qy 252 EIKTETTRFYVDYD-----LMLSLKKAQKMINTCN 286
Db 248 EKEMEINEXKLKTNDRSELETYELKVLRLVLEEVNESDRHLDTCN 294

RESULT 38
G97236
ATPase involved in DNA repair [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: G97236
R;Molling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97236
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1163 <KUR>
A;Cross-references: UNIPROT:Q97FKI; GB:AE001437; PIDN:AAK80682.1; PID:gl5025772; GSPDB:G
C;Genetics:
A;Gene: CAC2736

Query Match 7.3%; Score 110; DB 2; Length 1163;
Best Local Similarity 17.9%; Pred. No. 38;
Matches 73; Conservative 69; Mismatches 114; Indels 152; Gaps 14;

Qy 1 MTEIVADKTVEVVKNAIETADG-----ALDLYNKYLQVTPWQTFDETIK 45
Db 109 IVDITGDE-VEVLEGAQSVNEKQEIIGLSLDFTRTV--VLPQGFSEFLKLEKERR 165
Qy 46 -----ELSRFKQYSQAASVLVGDIK-----TLMDSQ 73

Db 166 NMLERLNLQYGDLSFKLARKIRKEREKENVLGKGVENINEDVLKERRELLKNN 225
Qy 74 DKYFEATQTVYEWCGVATQALAAAYILLFDBYN-----EKASQAQKILIKV-- 119
Db 226 DFFNEASK-----BYLKAEBEYNEGKEVWGLQIEIEEKRNVRKOLMEKDE 271
Qy 120 --LDDGTTKLNKAAQKSLVSSQSFNNASGKLLALDSQLTN-----DFSEKSS 164
Db 272 IDLKEKRLGESSKVKPYDNTYNTLTKQIDILKEQILSRENTMKAISLEKEMEKLS 331
Qy 165 YFQSQVDKIRKEAYAGAAAGVAGVPGFLIISYIAAGVVEGKIPKLNKLSVQNFTT 224
Db 332 IAKDNKEK-----ALPKFMIKHIIILDAIKEDKLLDNKLEKRLQGGKIEK 377
Qy 225 LS-----NTVQAKNDIDAQKLUKTEIAAAGEIKTETE----- 258
Db 378 LSLEASNKEELIKQNTKIDSLTLKIQNLSEKIDNLKVPPEYKKNKINIGIFLLRNYDEKL 437
Qy 259 -----TTRFVVDYD-----DLMLSLKKAQKMINTCNEYQKR 291
Db 438 KHKNKLGDCDKFQVDFEKAQSKKEMLFNKLEBERSKL-----DTYTKK 481

RESULT 39
JC5837
364K Golgi complex-associated protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C;Accession: JC5837
R;Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehata, Y.
Cell Struct. Funct. 22, 565-577, 1997
A;Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
F;49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F;3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 7.3%; Score 110; DB 2; Length 3187;
Best Local Similarity 20.8%; Pred. No. 1.4e+02;
Matches 80; Conservative 65; Mismatches 114; Indels 126; Gaps 18;

Qy 1 MTEIVADKTVE---VVKNA-IETADGALDLYNKYLQVTPWQTFDETIKLSRFKQYSQ 56
Db 1725 VTEAVVCKSQESLSSENKLEDAEATL-LANSAKPGVS--ETP-SSHDDINNYLQQLDQ 1780
Qy 57 AASVLVGDITKLLMDSQDKYFEATQTVYEWCGVATQALAAAYILLFDEYNEKKA----SAQ 112
Db 1781 ----LKGRIAELEMEKQ-KDELSQTL-----ENEKNALLTQISA 1815
Qy 113 KDILIKVLDDGITKLN-----EAQKSLVSSQSFNNASGKLLALDSQLTNDPSEKS-----S 164
Db 1816 KDSLEKLLBEVAKINMLNQIQEELSRTVKLTETABEEDKDLLEERLMNLAELNSIGN 1875
Qy 165 YFQSQVD-----KIRKAYAGAAAGV 186
Db 1876 YYQDVDTAQIKNQLSEMQNLKRCVSELEBEEKQQLVKEKTVKESEIRKE-YMEKIQAGQ 1934
Qy 187 AGPFGILISYSIAAGVVEGKILPELKNKLSVQ-----NFFTLSNTVK-----QA 232
Db 1935 KGFGSKIHAKELQ-----ELLKKEQEVKQLQKDCIRYLGRISALEKTVKALBFVHTES 1988
Qy 233 NKDIDAAK-----LKLTTETIAAIGIKETETETTRFYVDYDML----- 271
Db 1989 QKLDATKGNLAQAVEHHKKAQAEKLSFKILLDDTDEAARVLADNLKJLKEKLSQNKESI 2048
Qy 272 -SLLKEAAKMINTCNEYQKRGHK 295

Db 2049 KSOIKQKDEDLRLLEQAEKHKKE 2073

RESULT 40
T13216
minor capsid protein 1608 - Lactobacillus phage phi-gle
C;Species: Lactobacillus phage phi-gle
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: T13216
R;Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hirakawa, M.; Yamada, K.; Taketo,
Gene 187, 45-53, 1997
A;Title: Genome structure of the Lactobacillus temperate phage phi gle: the whole genome
A;Reference number: Z17631; MUID: 97225795; PMID: 9073065
A;Accession: T13216
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1608 <KOD>
A;Cross-references: EMBL: X98106; NID: e917136; PID: e247162; PIDN: CAA66745.1
C;Genetics:
A;Note: Rozf1608

Query Match 7.3%; Score 109.5; DB 2; Length 1608;
Best Local Similarity 19.6%; Pred. No. 62;
Matches 76; Conservative 67; Mismatches 137; Indels 129; Gaps 16;

Qy 1 MTEIVADKTVVEVKNAIETADGALDLYNKYLDQVWPQTFDETTELKLSRFKQEYSQAASV 60
Db 412 LVELI--KYVNAHNTIVDIIGNLKGILGIKTV-NKTFSDIVYDIAMK-----FG 460
Qy 61 LVGDIKTLMDSDQKYPEATQTVVEMCGVATQLLAAVILLFD-----EYNEKKAQAQDI 115
Db 461 LVGEKAQESDPLDKIDDALNLSKNOELIENLTAFIAMFALKKMGFEIGMLASLRKS 520
Qy 116 L-----TKVLD-----DGTITLNEAQSLLVSSQFNAS-----GKLLALDSQLTND 158
Db 521 IETAAVSKVDLFGSGVTSAGGKAVTQTVAKEAGGTAATAGSSKVLGRFAKGATSTA 580
Qy 159 FSEKSS-----YFQSQVDKIRKE-----AYA 179
Db 581 ELEAASGLGGKAMMAARGLTKAVPYMSIAASIPELFGTQTLGKHLGGFAGSAGGPAA 640
Qy 180 GAAA-----GVVAGPGLIISYIAAGVVEG--KLPELKNKLSV-----218
Db 641 GAAAGSAMPVYVAGVGIVGLAGSLQGVSGSIQKGTSPKLTSTKMSDLGHDMAK 700
Qy 219 -----QNFPTLSNTVKQANKDIDAALKLTTEIAAIGIKTETETT-----260
Db 701 KPSGSPKPSLNDKQFSKSYSLTKLNKQ---AKIKITDTSGISKAQKLDITTYGKM 757
Qy 261 -----RFY-----VDYDDLML--SLKBAKMKMINTCNEYQKRHK 294
Db 758 KKSVDKYGHKRWMSIKDYATLVQNGSMTEKEANKLNKAKENYKQAK 806

RESULT 41
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: A64465
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rsen, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID: 96337999; PMID: 8688087
A;Accession: A64465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1005 <BUL>
A;Cross-references: UNIPROT: Q58718; GB: U67572; GB: L77117; NID: g1591958; PIDN: AAB99331.1;
C;Genetics:

A;Map position: REV1273394-1270377
C;Superfamily: hypothetical protein MJ1322

Query Match 7.2%; Score 109; DB 2; Length 1005;
Best Local Similarity 19.1%; Pred. No. 37;
Matches 58; Conservative 55; Mismatches 93; Indels 98; Gaps 11;

Qy 7 DKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDE-----TIKLSRF-----KQEYSQ 56
Db 605 DEILEDIKSQLNKFK--NPNQYLSAVSYLSNVDSEGINRIKEIENIVSGWKKKRE 661
Qy 57 AASVL-----VGDITKLLMDSDQKYPEATQTVVEMCGVATQLLAAVILLFDVYNEKAS 110
Db 662 ELNKLREDEREINRLKDLNELKNKE-----LIEIENRRSRUKFDKYKEYLGL 710
Qy 111 AQDKIILKVLDDGITKLINEAQAQSLVSSQFNASGKLLALD-----152
Db 711 TEKLEELKNIKDGLSEI-----YINCNSKILLAIIDNKRKYNKEDIEIYLNK 757
Qy 153 -----SQTNDSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSTAAGVVEGKLI 208
Db 758 ILEVNKEINDIEERISVINOGLDE-----INYN-----EEHKKI 792
Qy 209 PEL-----KNKLKSVQNFPTLSNTVKQANKDIDAALKLTTEIAAIGIKTETETTFYV 264
Db 793 KELYENKROELDNVRBQKTEIETGIEYLKQVESLKRLL-----KEMSNLEKEKEKLTQFV 848
Qy 265 DYDD 268
Db 849 EYLD 852

RESULT 42
S54091
hypothetical protein YPR070w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YP9499.25
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C;Accession: S54091; S69058
R;Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54059
A;Accession: S54091
A;Molecule type: DNA
A;Residues: 1-566 <BAD>
A;Cross-references: UNIPROT: Q12321; EMBL: Z49219; NID: g805025; PID: g805050; MIPS: YPR070w
A;Experimental source: strain AB972
R;Couch, J.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of S. cerevisiae cosmid 9513.
A;Reference number: S69057
A;Accession: S69058
A;Molecule type: DNA
A;Residues: 1-566 <COU>
A;Cross-references: EMBL: U51033; NID: g1230676; PID: g1230678; MIPS: YPR070w
C;Genetics:
A;Gene: SGD:MED1
A;Cross-references: SGD: S0006274
A;Map position: 16R

Query Match 7.2%; Score 108.5; DB 2; Length 566;
Best Local Similarity 23.9%; Pred. No. 19;
Matches 78; Conservative 43; Mismatches 131; Indels 75; Gaps 18;

Qy 17 IETADGALDLYNKY-----LDQVWPQTFDETI-KELSRFKQEYSQAASVLV- 62
Db 8 VETLDSMIELFKDYKPGSITLENITRCQTLGLESFTELSNELSL-----STASKIIV 63
Qy 63 -----GDIKTLMDSDQK--YFEATQTVVEMCGVATQLLAAVILLFDVYNEKKA 109
Db 64 DVDYNNKQDRIDQVKVLASNFONFDYFNQDGEHEKSNILLNSLTYPDLKAFHNLKF 123
Qy 110 SAQDKIILKVLDDGITKLN-EAQSLLVSSQFNASGKLLALDSQLTNDPSEKSSY--- 165

Db 124 LYLDAYSHIESOSTSHNNGSSDKLSDSNASFN--QGLU-----DLFKYFTELUSHVIRQ 177
Qy 166 -FOSQV--DKIRKEAYAGAAAGVAGPGLIISYIAAGVVEGKLIPELK-----NKLKS 217
Db 178 CFQDNCCDFKVRTN-----LNDKFGI--YILTQG-INGKEVPLAKIYLEENKSDS 224
Qy 218 VQNFYFTL--SNTVKQANKDID--AAKLKLTTEIAAIGEIKTETETTRFYVDY----- 266
Db 225 QYRFYEIYISQETKSWINESAENFSGISLWMEIVANAKESNYTDLIWFPEDFISPELII 284
Qy 267 DDLMLSLKKAANKMIN--TCNEYOKR 291
Db 285 DKVTCSSNSSSPPIDLFNNYNSR 311

RESULT 43
H69378
conserved hypothetical protein AF1032 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: H69378
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleckmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: H69378
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-886 <KLE>
A;Cross-references: UNIPROT:O29230; GB:AE001032; GB:AE000782; NID:G2689355; PIDN:AA89021
A;Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 7.2%; Score 108.5; DB 2; Length 886;
Best Local Similarity 21.0%; Pred. No. 33;
Matches 74; Conservative 59; Mismatches 109; Indels 111; Gaps 15;
Qy 8 KTVVVKNAIETADGALDLYNKYLQVVPWQTFDETIKELSRFKQBYSQASVLVGDIKT 67
Db 425 KAVEELKSARTC-----PVCGRLEDEHRKNIMAYETREMKRIAELAKA-----DEIK 475
Qy 68 LL---MDSQDKYFEATQTVYEWCGVATQLLAAYITLLFDEYNKASAKQDILIKVLD--D 122
Db 476 KLERLEKVEKALEKQETLVKRYQWVDELKALENEL--SSHDAEKLKSAESEYRKVKERLD 534
Qy 123 GITKLNDAQKSLVSSOSFNNAAGKLLALDQLTNDSEKSSYFQSQVDKIRKEAYAGAA 182
Db 535 G---LRGQOKILLASSARIKELKSSLSREIBALKNVESERGELHR-----KIREEGF--- 583
Qy 183 AGVVAGPGLIISYIAAGVVEGKLIPELANKLKSQVNF-----TTLSNTVKOA 232
Db 584 -----ESLEBEREVQSLRPFFYNKWLKDAESRLESELEKRR 620
Qy 233 NK---DIDAAKLKL-----TTEIAAIGEI 253
Db 621 EKLEDEISEIAKLEANGKAEIRGQIDELLRIYSEEHRLSRDEHLRKSKELAG---L 677
Qy 254 KTTETTRFYVDYDMLSLKKAAGKQVINTCNEYQKRGHCKTLFE---VPEV 303
Db 678 KSRLETRESLQSAEKDLKFELEQAKW---DEYRK---KVEVFEXIAIPEL 723

RESULT 44
T44802
type I site-specific deoxyribonuclease (EC 3.1.21.3) chain hsdR [imported] - Streptococ
N;Alternate names: type IC restriction subunit
C;Species: Streptococcus thermophilus
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44802

R;Solow, B.T.; Somkuti, G.A.
submitted to the EMBL Data Library, August 1999
A;Description: Comparison of low molecular weight heat stress proteins encoded on plasmid
A;Reference number: Z22847
A;Accession: T44802
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1025 <SOL>
A;Cross-references: UNIPROT:Q9RNW2; EMBL:AF177167; PIDN:AAF04356.1
A;Experimental source: strain ST135
C;Genetics:
A;Gene: hsdR
A;Genome: plasmid pER35
C;Superfamily: type I site-specific deoxyribonuclease chain hsdR
C;Keywords: hydrolase

Query Match 7.2%; Score 108.5; DB 2; Length 1025;
Best Local Similarity 20.1%; Pred. No. 40;
Matches 80; Conservative 52; Mismatches 109; Indels 157; Gaps 19;
Qy 7 DKTVEVVKNAIETADGALDLYNKYLQVVPWQTFDETI-----KELSRFKQBYSQAS- 59
Db 628 DELVEIMKE-----YEQFD-ASPYQ--DEKLYNQNKRLARKEKQYQSDGQW 673
Qy 60 ---VLVCD-----IKTLLMDSQDKY-----FEATQTVYEWCGVATQLLAAY--- 97
Db 674 LDFVIVVDRLLTGFDSPTIQTLYIDREMNQYKLLQAFSRNRIY--TGKDSGLIVSRKP 731
Qy 98 -----ILLFDEYNE--KASAKQDILIKVLDGDTIKLNEAQSL 134
Db 732 FTVRENVRNTRFLPSNEKQNFQDLIPKEVEEVKKEFECSTLYKQSEADLSNDNDLKT 791
Qy 135 LVSSQSFNNAAGKLLALDS--QLTNDPSEKSSY----- 165
Db 792 IAQVSAVQKLEKSYKALRSYDQVEEDFEFSEVVEQLPQYQGTENIKTKIKEMIEDEGH 851
Qy 166 -----FQSQVDKIRKEAYAGAAAGVAGVGFGLIISYI-----NAGV 202
Db 852 PESDFEKLQEIAPSSQLNATHKD-----VDSFYINQLLKAQLNEAGA 896
Qy 203 VEGKLIPELANKLKSQVNFPTLSNTVKQANKDIDAAKLKLT---EIAAIGEIKTETET 259
Db 897 VE-KFEKSIQKDPQIQKMYHTLKDQLVNTEIDVAQLKETSIONEI---QRQLQKEA 951
Qy 260 TRFYVDYDMLSLKKAAGKQVINTCNEYQKRGHCKTL 297
Db 952 EEFGLSFDPLQSAM-----NEYQS--DKKTI 975

RESULT 45
A90551
conserved hypothetical protein MYPV 3130 [imported] - Mycoplasma pulmonis (strain UAB CT
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: A90551
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: A90551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2819 <KUR>
A;Cross-references: UNIPROT:Q98QPS; GB:AL445566; PID:G14089727; PIDN:CAC13486.1; GSPDB:G
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPV 3130
A;Genetic code: SGC3

Query Match 7.2%; Score 108.5; DB 2; Length 2819;
Best Local Similarity 21.3%; Pred. No. 1.5e+02;
Matches 77; Conservative 55; Mismatches 119; Indels 111; Gaps 20;

Query Match 7.2%; Score 108; DB 2; Length 2253;
Best Local Similarity 19.9%; Pred. No. 1.2e+02;
Matches 69; Conservative 60; Mismatches 146; Indels 72; Gaps 10;

Qy 6 ADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFD-ETIKELSRFKQEQSQAASVILVG 64
Db 1132 AEKEMAKLKSVISEKSRIFC-----LEQDTONKRDLSCTIOHQSKLGESQGLQALID 1187
Qy 65 I-----KTLMDSDQKYPEATQTVYEWCGVATQLLAAYILLDFEYNEKKASA-----Q 112
Db 1188 LEKKCKEQKELICEAQNKAAEAKTASEKASVSRQLEGIQALEIEIGKERKACDLQKQ 1247
Qy 113 KDILIKVLDGIIKLINEAQSKLL-----VSQSFNNAASKLLALDS-----QLTNDF 159
Db 1248 LELSWAVQESKETELQAKKELPHKVOELQSQTSFTDSSGEALLYLSEAQERQQALTEA 1307
Qy 160 SEKSPQSQSDVKIRKEAYAGAAAGVAGPFGILISYIAAGV-----VSGKLIPELKNK 214
Db 1308 KEQAEQYQKIEKMKNEVNSIQAE-----IKILSKVITNEEVSVDPEQLLKETS 1360
Qy 215 LKSVQNFPTLSNTVKQAKDIDAAKLKLTETAAGIEIKTETTRFRFYVDYDDLMLSL 274
Db 1361 AK-----LEEKQKHLMEASFPELLEKKNCAIDCLTTEAQNLKGEADQORMAVDSL 1412
Qy 275 -----KEAAKMINTCNEYKRGHKKTLE 299
Db 1413 QOKLSSKAETNHTLQELQIAQWKNCAEKEQICSLQNLKSNQSLLE 1459

RESULT 49
T18281
hypothetical protein D1 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18281
R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh
Genetics 148, 1117-1125, 1998
A:Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1
A:Reference number: Z14684; MUID:98198836; PMID:9539429
A:Accession: T18281
A:Status: preliminary; translated from GE/EMBL/DBDB
A:Molecule type: DNA
A:Residues: 1-1474 <R1E>
A:Cross-references: UNIPROT:Q23870; EMBL:U00796; NID:g2702254; PID:g2702256; PIDN:AAC186
C:Genetics:
A:Introns: 33/2

Query Match	7.1%; Score 107; DB 2; Length 1474;
Best Local Similarity	22.8%; Pred. No. 79;
Matches	79; Conservative 48; Mismatches 129; Indels 90; Gaps 16;
Qy	18 ETADGALDL-----YNKYLDQVWPQTFDETIKELSRFKQBYSOAA\$VL--VGDIKT 67
Db	777 ESQKGSNDIEIQKLANQSKYISTI---NDRDSTIKLQALINELNSIFKLNQOQSIKD 833
Qy	68 LLMDSDQKYFEATQVYECWGVATQLLAAAYILLFDEYNEKKASAQ--KDILIKYLDLDDGIT 125
Db	834 TLFNSAQLLTEKQKSNH---NMWLEKSEQEYKILLDEQIEKKNMANNVSEIKTKDEFTIE 890
Qy	126 KLNBAQKSLIVSSQSFNNASGKL-LALDSQ---LTNDFSEKSSYF-----SQQVDKI- 173
Db	891 VLNQTLNCTNESNTTIETYKKLHLENKTAILLNEISSQLYFDITISGYQYVINEYE 950
Qy	174 -----RKEAYAGAAAGVAGPPGLIITSYSTAAGVWEGKLIPELKNLKSQVONFTT--TL 225
Db	951 NCFSEKEKE-----LOVSY--AAGRVLKDKNDQINAELETLLKNNFKDFNL 993
Qy	226 SNTVYKQANK-----DIDAAKLJLTTEIAAIGETKTETET-----RFYVVDYD-- 267
Db	994 LNSUKISNEHKSQNLNLTNKNYSLEKTESLURSIIOLETTPTVSNQITOPAFYSYKHE 1053
Qy	268 ----DLMLSLI\$KEAAK-----KMINTCNEYQKRGHKKT 296

[illegible]

Search completed: January 5, 2005, 10:59:47
Job time : 21.4424 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:42:38 ; Search time 63.7089 Seconds
(without alignments)
1717.381 Million cell updates/sec

Title: US-09-993-292B-2

Perfect score: 1515

Sequence: 1 MTSIFAEQTVVVKSALETA.....NEYQRRGKKTLEVPDVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : A: Geneseqp23Sep04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1515	100.0	305	6	ABG73265 Salmonell
2	1515	100.0	749	6	ABG73266 Salmonell
3	1390	91.7	309	2	AAR67645 Haemolysi
4	1079	71.2	1035	4	ABG25221 Novel hum
5	121.5	8.0	1036	7	ADF07451 Bacterial
6	118.5	7.8	2681	6	ABJ19025 Pathogen
7	117.5	7.8	582	6	ABM68103 Phototrab
8	114.5	7.6	588	4	ABM71280 Drosophil
9	114	7.5	515	7	ADC95690 E. faeciu
10	112.5	7.4	1948	5	ABP73774 Candida a
11	112	7.4	1189	6	ABU42996 Protein e
12	111.5	7.4	1158	6	ABU28937 Protein e
13	111	7.3	941	6	ABU18419 Protein e
14	111	7.3	1231	7	ADC95523 E. faeciu
15	110.5	7.3	476	2	AAR43563 Hyalurona
16	110	7.3	951	4	ABU53070 Intracell
17	110	7.3	953	4	ABU53071 Intracell
18	110	7.3	961	4	ABU53077 Intracell
19	110	7.3	2228	7	ABR61599 Human gol
20	110	7.3	2230	6	ABU07445 Protein d
21	110	7.3	2230	7	ABR61600 Human gol
22	110	7.3	2250	7	ABR61601 Human gol
23	110	7.3	2252	7	ABR61602 Human gol
24	109.5	7.2	742	6	ABU25599 Protein e
25	109	7.2	739	4	AAB96493 Putative

26	109	7.2	929	4	ABU53079 Intracell
27	108.5	7.2	1048	4	ABBS9245 Drosophil
28	108.5	7.2	1066	4	AAG67418 Amino aci
29	108.5	7.2	1066	6	ABG72693 Fruitfly
30	108.5	7.2	1066	7	ADG98862 Fruit fly
31	108.5	7.2	3692	6	ABU43311 Protein e
32	108.5	7.2	3696	5	ABP40235 Staphyloc
33	107.5	7.1	1292	5	ABB77986 Amino aci
34	107.5	7.1	1875	6	ABR53560 Protein s
35	107.5	7.1	1875	7	ADK64380 Disease t
36	107	7.1	1211	5	ABP39975 Staphyloc
37	106.5	7.0	927	5	ABP39975 Staphyloc
38	106.5	7.0	1312	2	AAW22775 Human RAD
39	106.5	7.0	1312	2	AAW71295 Human hom
40	106.5	7.0	2346	4	ABB63519 Drosophil
41	106.5	7.0	4734	8	ADH62807 Lactobaci
42	106.5	7.0	7201	4	ABB71136 Drosophil
43	106	7.0	685	4	ABB64018 Drosophil
44	106	7.0	815	6	ADA32997 Acinetoba
45	106	7.0	863	6	ADB08930 Alloiococ
46	106	7.0	885	4	AGS82283 S. epider
47	106	7.0	2835	5	AAU75100 Novel sig
48	105.5	7.0	1318	5	ABB77985 Amino aci
49	105.5	7.0	1318	7	ADJ68860 Human hea
50	105.5	7.0	1318	8	ADJ66509 RAD50 hom

ALIGNMENTS

RESULT 1

ABG73265

ID ABG73265 standard; protein; 305 AA.

XX ABG73265;

XX 17-APR-2003 (first entry)

XX Salmonella typhi cytolysin A (ClyA) protein.

DE Protein expression vector; expression cassette; export protein;
XX transformed host cell; bacterial cell; immune response; HlyE family;
KW cytolysin A; ClyA; recombinant protein; bioreactor; bacterial growth;
KW live vector immunogenic composition; animal feed; animal vaccination;
KW food industry; nutritional supplement; biomediation; waste disposal;
KW waste treatment; host bacterium.

OS Salmonella typhi.

XX US2002146430-A1.

XX 10-OCT-2002.

XX 23-NOV-2001; 2001US-00993292.

XX 22-NOV-2000; 2000US-0252516P.

XX (GALE/J) GALEN J E.

PI Galen JE;

XX WPI; 2003-228013/22.

DR N-PSDB; ABX15174.

XX Novel system for expressing protein of interest, has expression vector
PT with cassette having export protein coding sequence fused to protein
PT coding sequence, host cell transformed with vector, and culturing
PT environment.

XX Claim 5; Page 18; 33pp; English.

XX The present invention relates to a system for expressing a protein of
CC interest. The system comprises an expression vector having an expression

CC cassette, where the expression cassette comprises an export protein
 CC coding sequence genetically fused to a DNA sequence encoding the protein
 CC of interest, a host cell transformed with the expression vector, and a
 CC culturing environment for the transformed host cell, where the expression
 CC cassette expresses the export protein-protein of interest fusion protein
 CC which is exported out of the host cell. The system is useful for
 CC expressing a gene in a bacterial cell, by providing an expression vector
 CC to a population of untransformed bacterial cells, and expressing the
 CC expression cassette, such that the export protein-protein of interest
 CC fusion protein is produced and exported into the culture medium. The
 CC protein of interest is preferably an antigen and this method is useful
 CC for eliciting an immune response from a host against the fusion protein.
 CC The export protein may be a member of the HlyE family such as cytolyisin A
 CC (ClyA). The system is useful for the production of recombinant proteins
 CC inside a bacterial host cell, in a bioreactor, and various devices that
 CC facilitate the growth of bacteria. The system is also useful for
 CC providing an animal antigenic material against which an immune response
 CC may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can
 CC be used as a live vector immunogenic composition capable of facilitating
 CC the generation of an immune response in an animal. The system is also
 CC useful for preparing immunogenic compositions for vaccinating animals,
 CC and for use in the food, nutritional supplement, animal feed,
 CC biomediation, waste disposal, and waste treatment industries. The system
 CC is very advantageous for protein expression, as proteins that are toxic
 CC to host bacterium can also be expressed. A population of recombinant host
 CC cells can be transformed with the expression vector, and then the
 CC population of bacterial host cells can be maintained in culture and used
 CC to produce proteins without having to harvest and lyse the host cells.
 CC The present represents Salmonella typhi ClyA protein

XX Sequence 305 AA;

Query Match 100.0%; Score 1515; DB 6; Length 305;
 Best Local Similarity 100.0%; Pred. No. 9.5e-116;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSIFABQTVVEVKSATETADGALDLYNKYLDQVTPWKTDETTIKELSRFKQEYSQASV 60
 Db 1 MTSIFABQTVVEVKSATETADGALDLYNKYLDQVTPWKTDETTIKELSRFKQEYSQASV 60

Qy 61 LVGDIKVLMDSDQKPYEATQTVYEWGVTQLLSAVILLFDEYNEKKAQKDILIRIL 120
 Db 61 LVGDIKVLMDSDQKPYEATQTVYEWGVTQLLSAVILLFDEYNEKKAQKDILIRIL 120

Qy 121 DDGVKLNKAQKSLTSSQSFNNAAGKLLALDSQLTNDFSEKSSYFQSDVDRKEAYAG 180
 Db 121 DDGVKLNKAQKSLTSSQSFNNAAGKLLALDSQLTNDFSEKSSYFQSDVDRKEAYAG 180

Qy 181 AAAGIVAGPGLIITSYISIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
 Db 181 AAAGIVAGPGLIITSYISIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240

Qy 241 LKLAETAAAGIEIKTETTRFYVDYDMLSLKGAQKXWINTCNEYQQRHGKTLFEV 300
 Db 241 LKLAETAAAGIEIKTETTRFYVDYDMLSLKGAQKXWINTCNEYQQRHGKTLFEV 300

Qy 301 PDVAS 305
 Db 301 PDVAS 305

RESULT 2

ABG73266

ID ABG73266 standard; protein; 749 AA.

XX AC ABG73266;

XX DT 17-APR-2003 (first entry)

XX Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.

DE Protein expression vector; expression cassette; export protein;

XX transformed host cell; bacterial cell; immune response; HlyE family;

KW

KW cytolyisin A; ClyA; recombinant protein; bioreactor; bacterial growth;
 KW live vector immunogenic composition; animal feed; animal vaccination;
 KW food industry; nutritional supplement; biomediation; waste disposal;
 KW waste treatment; host bacterium; sacB; mutant; mutein;
 KW lethal levansucrase.

XX Salmonella typhi.

OS Bacillus subtilis.

OS Synthetic.

OS Chimeric.

XX US2002146430-A1.

XX 10-OCT-2002.

XX 23-NOV-2001; 2001US-00993292.

XX 22-NOV-2000; 2000US-0252516P.

XX (GALE/) GALEN J E.

XX Galen JE;

XX WPI; 2003-228013/22.

XX N-PSDB; ABX15191.

XX Novel system for expressing protein of interest, has expression vector

XX with cassette having export protein coding sequence fused to protein

XX coding sequence, host cell transformed with vector, and culturing

XX environment.

XX Example 3; Page 25-28; 33pp; English.

XX The present invention relates to a system for expressing a protein of
 XX interest. The system comprises an expression vector having an expression
 XX cassette, where the expression cassette comprises an export protein
 XX coding sequence genetically fused to a DNA sequence encoding the protein
 XX of interest, a host cell transformed with the expression vector, and a
 XX culturing environment for the transformed host cell, where the expression
 XX cassette expresses the export protein-protein of interest fusion protein
 XX which is exported out of the host cell. The system is useful for
 XX expressing a gene in a bacterial cell, by providing an expression vector
 XX to a population of untransformed bacterial cells, and expressing the
 XX expression cassette, such that the export protein-protein of interest
 XX fusion protein is produced and exported into the culture medium. The
 XX protein of interest is preferably an antigen and this method is useful
 XX for eliciting an immune response from a host against the fusion protein.
 XX The export protein may be a member of the HlyE family such as cytolyisin A
 XX (ClyA). The system is useful for the production of recombinant proteins
 XX inside a bacterial host cell, in a bioreactor, and various devices that
 XX facilitate the growth of bacteria. The system is also useful for
 XX providing an animal antigenic material against which an immune response
 XX may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can
 XX be used as a live vector immunogenic composition capable of facilitating
 XX the generation of an immune response in an animal. The system is also
 XX useful for preparing immunogenic compositions for vaccinating animals,
 XX and for use in the food, nutritional supplement, animal feed,
 XX biomediation, waste disposal, and waste treatment industries. The system
 XX is very advantageous for protein expression, as proteins that are toxic
 XX to host bacterium can also be expressed. A population of recombinant host
 XX cells can be transformed with the expression vector, and then the
 XX population of bacterial host cells can be maintained in culture and used
 XX to produce proteins without having to harvest and lyse the host cells.
 XX The present represents Salmonella typhi ClyA protein

XX Sequence 749 AA;

Query Match 100.0%; Score 1515; DB 6; Length 749;

Best Local Similarity 100.0%; Pred. No. 3e-115;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVVKSAIETADGALDLNKKYLDQVIPWKTFDETIKELSRFKQEYSQASV 60
 DB 1 MTSIFAEQTVVVKSAIETADGALDLNKKYLDQVIPWKTFDETIKELSRFKQEYSQASV 60
 QY 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120
 DB 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120
 QY 121 DDGVKKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIRKEAYAG 180
 DB 121 DDGVKKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIRKEAYAG 180
 QY 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
 DB 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
 QY 241 LKLAETAAIGEIKTETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTUFEV 300
 DB 241 LKLAETAAIGEIKTETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTUFEV 300
 QY 301 PDVAS 305
 DB 301 PDVAS 305

RESULT 3

AA67645
 ID AAR67645 standard; protein; 309 AA.
 AC AAR67645;
 XX
 XX 25-MAR-2003 (revised)
 DT 06-JUL-1995 (first entry)
 XX
 XX Haemolysin regulator.
 XX Haemolysin regulator; tuberculosis; vaccine; therapy.
 XX Escherichia coli.
 XX WO9428137-A1.
 XX
 XX 08-DEC-1994.
 XX
 XX 24-MAY-1994; 94WO-US005869.
 XX
 XX 24-MAY-1993; 93US-00066830.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX King CH, Shinnick TM, Sathish M;
 XX
 XX WPI; 1995-022809/03.
 XX P-PSDB; AAQ75857.
 XX
 XX New nucleic acid encoding haemolytic regulator of E coli - and derived
 XX PT vectors, proteins etc used to transform Mycobacterium cells to produce
 XX PT antituberculosis vaccines providing improved immune response.
 XX
 XX Claim 9; Page 39-40; 46pp; English.
 XX
 XX The sequence represents the E. coli haemolysin regulator which may be
 XX used as an immunogen in a vaccine composition against tuberculosis
 XX CC (Mycobacterium tuberculosis). (Updated on 25-MAR-2003 to correct PN
 XX CC field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-
 XX CC 2003 to correct PA field.)
 XX
 XX SQ Sequence 309 AA;

Query Match 91.78; Score 1390; DB 2; Length 309;
 Best Local Similarity 91.08; Pred. No. 1.7e-105;
 Matches 274; Conservative 15; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVVKSAIETADGALDLNKKYLDQVIPWKTFDETIKELSRFKQEYSQASV 60
 DB 1 MTSIFAEQTVVVKSAIETADGALDLNKKYLDQVIPWKTFDETIKELSRFKQEYSQASV 60
 QY 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120
 DB 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDILIRIL 120
 QY 121 DDGVKKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIRKEAYAG 180
 DB 121 DDGVKKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSQVDRIRKEAYAG 180
 QY 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
 DB 181 AAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
 QY 241 LKLAETAAIGEIKTETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTUFEV 300
 DB 241 LKLAETAAIGEIKTETTRFYVDYDDLMLSLKGAACKMINTCNEYQORHGKKTUFEV 300
 QY 301 P 301
 DB 301 P 301
 RESULT 4
 ABG25221
 ID ABG25221 standard; protein; 1035 AA.
 AC ABG25221;
 XX
 XX 18-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #25212.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 XX
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR
 XX N-PSDB; AAS89408.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX PT diagnostics, forensics, gene mapping, identification of mutations
 XX PT responsible for genetic disorders or other traits and to assess
 XX PT biodiversity.
 XX
 XX Claim 20; SEQ ID NO 55580; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 XX sequences. (I) is useful as hybridisation probes, polymerase chain
 XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 XX CC and in recombinant production of (II). The polynucleotides are also used
 XX CC in diagnostics as expressed sequence tags for identifying expressed
 XX CC genes. (I) is useful in gene therapy techniques to restore normal
 XX CC activity of (II) or to treat disease states involving (II). (II) is
 XX CC useful for generating antibodies against it, detecting or quantitating a
 XX CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 1035 AA;

Query Match 71.2%; Score 1079; DB 4; Length 1035;
 Best Local Similarity 92.2%; Pred. No. 2.4e-79;
 Matches 212; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
 QY 70 MDSQKYFEATQTVYVCGVVTQLLSAVILLFDEYNEKKAQAOKDILIRILDDGVKKLNE 129
 DB 1 MDSQKYFEATQTVYVCGVATQLLAYILLFDEYNEKKAQAOKDILIKVLDGDTKLINE 60
 QY 130 AQSLLTSQSQFNNSAGKLLALDSQLTNDPSEKSSYFQSQVDRIKRAYAGAAAGIVAGP 189
 DB 61 AQSLLVSSQSQFNNSAGKLLALDSQLTNDPSEKSSYFQSQVDRIKRAYAGAAAGVAGP 120
 QY 190 FGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTLSATVQKANKDIDAAKILATEIAA 249
 DB 121 FGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTLSATVQKANKDIDAAKILATEIAA 180
 QY 250 IGEIKTETTRFYVDYDDMLSLKGAAKKMWNTCNEYQORHGKTLFE 299
 DB 181 IGEIKTETTRFYVDYDDMLSLKGAAKKMWNTCNEYQORHGKTLFE 230

RESULT 5

ADFO7451
 ID ADFO7451 standard; protein; 1036 AA.
 XX
 AC ADFO7451;
 XX
 DT 12-FEB-2004 (first entry)
 XX
 DE Bacterial polypeptide #3564.
 XX
 KW Proteus mirabilis infection; bacterial infection; antibacterial;
 XX immunostimulant.
 XX
 OS Proteus mirabilis.
 XX
 PN US6605709-B1.
 XX
 PD 12-AUG-2003.
 XX
 PF 05-APR-2000; 2000US-00543681.
 XX
 PR 09-APR-1999; 99US-0128706P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Breton GL;
 XX
 DR WPI: 2003-895291/82.
 DR N-PSDB; ADF03279.
 XX
 PT New Proteus mirabilis polypeptides and polynucleotides, useful as
 PT reagents for diagnosis of bacterial disease, as components of
 PT antibacterial vaccines, as targets for antibacterial drugs, or as
 PT biocontrol agents for plants.
 XX
 PS Disclosure; SEQ ID NO 7736; 870pp; English.

XX

CC The invention relates to new Proteus mirabilis polypeptides and
 CC polynucleotides. The invention also relates to antibodies against the
 CC polypeptides, methods for producing the polypeptides, a method of
 CC generating vaccines for immunising an individual against P. mirabilis, a
 CC method for evaluating a compound for the ability to bind a P. mirabilis
 CC polypeptide and a method for screening test compounds for anti-bacterial
 CC activity. The polypeptides and polynucleotides are useful as molecular
 CC targets for diagnosing, preventing and treating pathological conditions
 CC resulting from bacterial infection, as reagents for diagnosis of
 CC bacterial diseases, as components of antibacterial vaccines, as targets
 CC for antibacterial drugs or as bio-control agents for plants. This
 CC sequence represents a Proteus mirabilis polypeptide of the invention.

XX
 SQ Sequence 1036 AA;

Query Match 8.0%; Score 121.5; DB 7; Length 1036;
 Best Local Similarity 20.7%; Pred. No. 0.67;
 Matches 80; Conservative 66; Mismatches 126; Indels 115; Gaps 17;
 QY 2 TSIFAEQTVYVVK-----SAIFADGALDLYNKLQVTPWKTFTDIKELSRPKQE 53
 DB 167 TALSSMEINELIKKQNGENVSSSELAKASIDLINQLVD-----TASSLNNNISAFSQ 220
 QY 54 YSQEASVL-----VGDIKVLLMDSQKYFEATQTVYVCGVVTQLLSAVILL-FDEY 104
 DB 221 LNKLGSLVNTKHLNGVGN-KLQNLPNLDKUGTGLDTV---SGILSAISAFILSNADAD 276
 QY 105 NEKKAQAOKDILIRILDDGVKKLNE-----AQKSLTSSQSQFNNSAGKLL 149
 DB 277 TGTAAAGVELTQVILGNVKAQSVQYLAQRAVQAQGLSTSAAGSLTASVLAISPLSEL 336
 QY 150 ALDSQL--TNDPSEKSSYFQ-----SQVDRIKRA-----YAGAA 183
 DB 337 AIADQFRANKIKEYSQRFKFGYEGDLSLLAFRKETGAIDASLTINTALGTISAGISA 396
 QY 184 GIVAGPFGILISYIAA--GVIEG-----KLIPELNNRLKTV-----QNFETS- 224
 DB 397 ASTASLIGAPISALVGAITGIISGILEASKQSFHVHVRMANTIAEWEKTHGKNFFENG 456
 QY 225 -----LSATVK---QANKOIDAAKILAT-----EIAAIGEIKTETTRFYV 264
 DB 457 YDARHSAFTLEDNFKLLSQYNKEYSVERSVLITQOHWDDELIGELASVTKNGAKTSGKSYI 516
 QY 265 DYDDLMLSLKGAAKKMWNTCNEYQOR 291
 DB 517 DYVE-----EGKRLKPKNEFQKO 535

RESULT 6

ABU19025
 ID ABU19025 standard; protein; 2681 AA.
 XX
 AC ABU19025;
 XX
 DT 06-MAR-2003 (first entry)
 XX
 DE Pathogen specific antigen related staphylococcal protein SEQ ID NO 274.
 XX
 KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
 KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
 KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
 KW autoimmune disease; HIV; hepatitis.
 XX
 OS Staphylococcus sp.
 XX
 PN WO200259148-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 21-JAN-2002; 2002WO-EP000546.
 XX
 PR 26-JAN-2001; 2001AT-00000130.

(CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PA Meinke A, Nagy E, Von Absen U, Klade C, Henics T, Zauner W;
 PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
 PI Tempelmaier B;
 XX
 XX WPI; 2003-075410/07.
 XX
 XX Identifying, isolating and producing hyperimmune serum-reactive antigens
 PT from a pathogen, for preparing vaccine or medicament for treating or
 PT preventing e.g. staphylococcal infections, comprises providing antibody
 PT preparation.
 XX
 XX Example 7; Page 200; 252pp; English.
 XX
 XX The invention relates to a novel method for identifying, isolating and
 CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
 CC allergen, a tissue or host prone to auto-immunity, where the antigens are
 CC used in a vaccine, comprises providing antibody preparation from a plasma
 CC pool of a type of animal, or individual sera with antibodies against the
 CC specific pathogen, tumour, allergen, tissue or host prone to auto-
 CC immunity. The hyperimmune serum-reactive antigens comprising any of the
 CC 62 sequences of 53-2261 amino acids fully defined in the specification,
 CC or their hyperimmune fragments are useful for the manufacture of a
 CC pharmaceutical preparation, particularly a vaccine against staphylococcal
 CC infections or colonisation against *S. aureus* or *S. epidermidis*. The
 CC preparation of antibodies is useful for the manufacture of a medicament
 CC for treating or preventing staphylococcal infections or colonisation
 CC against *S. aureus* or *S. epidermidis*. The antibody preparations may also
 CC be used for diagnostic and imaging purposes. Other conditions that can be
 CC treated include cancer, autoimmune diseases or infections caused by viral
 CC (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This
 CC sequence represents a staphylococcal protein relating to the method for
 CC identifying and producing pathogen specific antigens of the invention
 XX
 SQ Sequence 2681 AA;
 Query Match 7.8%; Score 118.5; DB 6; Length 2681;
 Best Local Similarity 21.3%; Pred. No. 4;
 Matches 72; Conservative 50; Mismatches 125; Indels 91; Gaps 13;
 QY 2 TSIPAEQVEVKSASIAETAGALDLYNKYLDVPIPKTFDET-----IKELSR 49
 DB 998 TDLKSEKLDLTKS-YKTLDTSFNDFVKYIDEMTMRIADETEKVNKKYDYLQKWLSD 1056
 QY 50 FKQRY-----SQEASVLVGDIKVL-----LMSDQKYPEATQTVYEW-C-G 88
 DB 1057 YMKKYDCILEISKYSNDAADKVLGDFTAITELQNDQFDVKNWAEFKQTLESFKDG 1116
 QY 89 VVTQLLSAYILLFDYNEKASAKQDILIRLDDGVKKLNEAQAOKSLTSSQSFNNASGKL 148
 DB 1117 IVT-----EAEKARLRVLQDMLDRESMD-----IEERYKSLLANQYTTNTDIKNRL 1161
 QY 149 LALDSQLTNDSEKSYFQSQVDRIKEAVAGAGIVAGPFGILISYSIAAGVIECKLI 208
 DB 1162 TA-----SRSPYLSVHAS-LRK-----VISEQIADGKVDSEK 1193
 QY 209 PELANRLKTVQNFPTSLSATVQKANKDI-----DAAKKLATEIAAIGIKTETETTRF 262
 DB 1194 TLANNSLNTYNTTUTATYSKTIQEBALNTLSQIISDVASKVKEENGVIITISSDVTIKK 1253
 QY 263 YVD-----YDDLMLSLKGAOKMINTCNEYQORHOK 294
 DB 1254 QRDGAVITYYSGVPTLSNDPAKSW--TTNDLKDLHLK 1289
 RESULT 7
 ABM68103
 ID ABM68103 standard; protein; 582 AA.
 XX
 XX AC ABM68103;
 XX
 XX DT 20-NOV-2003 (first entry)

Photorhabdus luminescens protein sequence #1200.

Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 detection; food; gene expression; plant; animal; microorganism; toxin;
 antibiotic; biopesticide; virulence factor; disease model; plague;
 whooping cough.

Photorhabdus luminescens.

WO200294867-A2.

28-NOV-2002.

07-FEB-2002; 2002WO-IB003040.

07-FEB-2001; 2001FR-00001659.

(INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.

Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 Buchrieser C;
 WPI; 2003-148459/14.

Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.

Claim 2; SEQ ID NO 1200; 1205pp; French.

The invention relates to the isolation of genes and their encoded
 proteins from Photorhabdus luminescens. The isolated sequences are
 sources of probes and primers for detecting the genome of *P. luminescens*
 and related species; to study polymorphisms; for gene analysis and for
 detection/amplification of the genes. Antibodies (Ab) raised against the
 polypeptides encoded by the genes are used for detection/identification
 of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that
 carry a gene-containing vector are used to select compounds that
 modulate, regulate, induce or inhibit expression of the genes in plants,
 animals or microorganisms other than *P. luminescens* and are able to alter
 response or sensitivity to toxins and antibiotics produced by *P.*
luminescens. Cells transformed to express the genes are useful for
 recombinant production of the proteins, particularly toxins and
 antibacterials useful as insecticides, bactericides and fungicides. The
 genes, proteins, vectors containing the genes and Ab are also useful
 therapeutically (to treat microbial infection by bacteria or fungi that
 are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
 biopesticides. Other uses of the genes and the proteins are as virulence
 factors and for identifying targets of human diseases for which *P.*
luminescens is a model (particularly plague and whooping cough). This
 sequence represents one of the isolated *P. luminescens* proteins

Query Match 7.8%; Score 117.5; DB 6; Length 582;
 Best Local Similarity 19.2%; Pred. No. 0.68;
 Matches 73; Conservative 56; Mismatches 142; Indels 109; Gaps 12;
 QY 17 IETADGALDLYNKYLDQV---IPWKTF-----DETI 44
 DB 133 LETPPPAYFLPYIIDQIKSWSPFNSFENLGQVSNWKGPIKYFTGYLKEPHDIEINI 192
 QY 45 KELSFRKQEYSQBASVLVGDIKVL-----SAQKDL-----LMSDQDK 75
 DB 193 YEYSEIKESAHKIEKFSQSAVEIVDNTVDSPTLNDNDFIKQNEIKNELYDLIDYQTR 252
 QY 76 YFEATQTVYEWGVTQLLSAYILLFDEYNEKA-----SAQKDL----- 116
 DB 253 LYDAQATI---TSNIYDLESQVALAITSANELEADYKFAVESIPTDYLECPLOCGLTHDNS 309
 QY 117 ----IRLLDGVKKLNEAQAOKSLTSSQSFNNASGKLALDSQLTNDSEKSYFQSQVDRI 172

PF 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
FI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA46866.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 70920; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC growth of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 1189 AA;
Query Match 7.4%; Score 112; DB 6; Length 1189;
Best Local Similarity 18.3%; Pred. No. 4.8;
Matches 61; Conservative 60; Mismatches 130; Indels 82; Gaps 9;
QY 7 EQTEVVVK-----SAIETADGALDLYNKLYDQVFPKTFDETKELSRFKQYQASV 60
D 163 EESAGVLKYKRAESIQKLDHTEDNLNVRVEDILYDLGRVPEKKEAAIKYQLSKE 222
QY 61 LVGDIKVLMDSDQKYEAQTQVVEWC-----GVVTLQLLSAYILLDFEYNE 106
D 223 MEQSDVIIVTSDIDHYEDNQRLDERLNLKLSQQAQEGQQAQINQLQRY----- 273
QY 107 KKAQAQKDIILRIIDGKVKLNQAQKSLTSSQFNWASCKLALDSQLTND-----PS 160
D 274 -GKRRQNN-----DYDIEKLN-----VELVKATENYQLSGKLVBERKKNQSETWATYE 323
QY 161 EKSSYFQSOVDRIKEAYAGAAGIVAGPFGLLIYSIAAGVIEGKLIPELNRL-----K 216

Db 324 EELDNLESQIDSINKKAQNE-----KLLADLNKKQQLNK 359
QY 217 TVQNFFTSLSATVKQANKDIDAAKLATRIAAIGIKTETTRFYDYDDMLSLKKG 276
D 360 EVOESELLVSDQHDQDEKLEEKNSYITLMSQSDVNDIRFLEHTINEAKKSLRDS 419
QY 277 AAKWMINTCNEYQQ-----RHGKKTLEFV 300
D 420 RLVEAFNLQKDIOQNTOTNKAYQSSKSKMQV 452
RESULT 12
ABU28937
ID ABU28937 standard; protein; 1158 AA.
XX AC ABU28937;
DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #14464.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Enterococcus faecalis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA32807.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 56861; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the

XX 01-JAN-2004 (first entry)
XX E. faecium protein sequence SEQ ID 5150.
XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
XX abdominal-pelvic infection.
XX Enterococcus faecium.
XX USG583275-B1.
XX 24-JUN-2003.
XX 30-JUN-1998; 98US-00107532.
XX 02-JUL-1997; 97US-0051571P.
XX 14-MAY-1998; 98US-0085598P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
XX WPI; 2003-799836/75.
XX N-PSDB; ADC91869.
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
XX Enterococcus faecium polypeptide useful for detection, prevention and
XX treatment of a pathological condition resulting from a bacterial
XX infection.
XX Example 1; SEQ ID NO 5150; 243pp; English.
XX The invention relates to an isolated nucleic acid derived from
XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having
XX one of 10 fully defined sequences given in the (or comprising 40
XX sequential nucleotides chosen from any of the nucleic acids, its
XX complement or sequences hybridising to it). Also included are a
XX recombinant vector comprising the nucleic acid operably linked to
XX transcription regulatory element, a cell comprising the vector and a
XX single-stranded probe comprising the nucleic acid. The nucleic acids are
XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
XX The nucleic acids is useful for diagnosing pathological conditions
XX resulting from E. faecium bacterial infection (e.g. urinary tract
XX infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
XX infection) and for screening drugs such as agonists and antagonists. The
XX nucleic acid is useful for recombinant production of Candida albicans -
XX derived peptides or antisense polypeptides. Pharmaceutical compositions
XX and vaccines containing the nucleic acid are useful for preventing or
XX treating Enterococcus faecium infections. The present sequence represents
XX one if the disclosed E. faecium proteins.
XX
XX Sequence 1231 AA;
Query Match 7.3%; Score 111; DB 7; Length 1231;
Best Local Similarity 23.3%; Pred. No. 6.1;
Matches 78; Conservative
6 AEQTVFVVKSAETADGALDLYNK-----YLDQVTPWKTDE-----TIKELS 48
269 AEEVIVSGVQAQTEGTIDYLVNANGKTGFLNIHLRYPPVETFFQKLPHVTKTIAVLID 328
49 RFKQEYSQASVLVDIKVLIMDSQ-----DKYFEATQVYEWGVVQTLLSAVILLF 101
329 RSKFPGAGGEPALL-DVQSALYDSBLRPVIGRGYGLGSK-----DVTDPQISA---VF 378
102 DEYNEKASAKQDILILIDGVKKLNEAKSL-LTSSQSFN-----NAGSKL 148
379 DELKKDPSIRKRTTIGVDDVTYQSLPKESLDLTPQTFOAKFWGFGSDGTGVANKSA 438
149 LALDSQLTNDFSKSSVFSQ-----VDRIR-----KEAYAGAAAGIVA--GPFGLII 194
439 IKITGDHTDKYAQGYFYDYSKSGSLTVSHLRFQDTPIRSAVLVEHADLVACTP-AYLH 497

QY 195 SYSIAGVIEGKLIPELNNRLKTVQFFTSLSATVQ--ANKDI---DAAKUKLATEIAA 249
DB 498 SYDLVKGKFGGIF--LLNTLWSDQLETHLPKLRXYLAENNIRFYTNAMRLAQEVGL 555
QY 250 IGEIKTETETTRF-----YVDYDDMLSLKGAACK 280
DB 556 GRRINTAMETAFFKLADIIPFDE-VLPFLKBEALK 589
RESULT 15
AAR43563
ID AAR43563 standard; protein; 476 AA.
XX
XX AAR43563;
AC
XX 25-MAR-2003 (revised)
DT 09-JAN-2003 (revised)
DT 05-APR-1994 (first entry)
XX
DE Hyaluronan receptor.
XX
XX Hyaluronan binding protein; HA; RHAMM; mediated motility; wound; healing;
KW diagnosis; treatment; cell locomotion; tumour invasion; birth defects;
KW inflammatory disorder; Alzheimer's disease; dementia;
KW Parkinson's diseases; Huntington's disease; AIDS; diabetes; auto;
KW immune diseases; corneal dysplasia; hypertrophy; surgery; burns; strokes;
KW multiple sclerosis; depression; schizophrenia; CNJ; contraception;
KW in vitro fertilisation; embryo development.
XX
OS Rattus sp.
XX
XX W09321312-A1.
PN
XX 28-OCT-1993.
PD
XX 13-APR-1993; 93WO-CA000158.
PF
XX 09-APR-1992; 92GB-00007949.
PR
XX (UYMA-) UNIV MANITOBA.
PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
PA
PI Turley EA;
XX
XX WPI; 1993-351722/44.
DR N-PSDB; AAQ51212.
XX
XX DNA encoding hyaluronan receptor - used to produce proteins and
XX antibodies for alteration of cell locomotion.
XX
XX Claim 7; Fig 23; 88pp; English.
XX
XX The sequence is that encoded by a cDNA clone encoding the hyaluronan
XX receptor (HARC). The sequence was obtd. by screening a 3T3 library in
XX lambda gt11 with antibodies to HARC. A clone of 1.9 kb was obtained and
XX used to rescreen the library to obtain the full length, 2.9 kb clone. HA
XX is down regulated in stationary normal cells and is only expressed in
XX situations where cell motility is desired, e.g. in wound healing, in
XX response to growth factors and in chemotaxis by white blood cells. HA may
XX be used for diagnosis and treatment of diseases involving cell
XX locomotion, e.g. tumour invasion, birth defects, acute and chronic
XX inflammatory disorders, Alzheimer's and other forms of dementia, AIDS,
XX diabetes, autoimmune diseases, corneal dysplasias and hypertrophies,
XX burns, surgical incisions and adhesions, strokes, multiple sclerosis,
XX depression/schizophrenia related to neuronal growth and pain states
XX involving nerve sprouting; also in CNJ and spinal cord regeneration,
XX contraception, in vitro fertilisation and embryo development. See also
XX AAR46548-51. (Updated on 09-JAN-2003 to add missing OS field.) (Updated
XX on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 476 AA;

Query Match 7.3%; Score 110.5; DB 2; Length 476;
Best Local Similarity 20.4%; Pred. No. 2;
Matches 68; Conservative 73; Mismatches 125; Indels 67; Gaps 18;
QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDQVWPWKTFDETIKELSRFKQYSEASV 60
DB 99 LDNLLREKEVELEKHIAHQQAIIIAQEKYIDTA--QSLRVVTAQLESQEKYNDTAQS 155
QY 61 LVGDIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKASQAQKIDILIRIL 120
DB 156 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 205
QY 121 DGGVKLNEAQKSLITSS---QSPNNAAGKLLALDLSOLTN-DFSEKSSYFQSQVDRIKE 176
DB 206 ESQVEKNDTAQSLRDVSAQLESYKSTLKEIE-DLKLNLTLQEKVMAEKSVEDVQQQ 264
QY 177 AYAGAAAGVAGPFGILIIYSIAAGVIEGKLIPELNRR-----LKTQVQNF--TSL 225
DB 265 ILTAESTN-----QEVA-----RMVQDLQNRSTLKEEIEKITSSFLEKITDL 307
QY 226 SATVQANKOI-----DAKCLKATEI-----AAGIKETETETTFYVD-YD 267
DB 308 KNQLRQODEFRKQLEKEGKRTAEKENVMTELTMEINKWALLYELVEKTRPFQOQLDAFE 367
QY 268 DLMLSLIK--GAAKMINTCNE-YQORHGKTL 297
DB 368 AEQALLNEHGATQEQQLNKIRDSYAQLLGHQNL 400

RESULT 16
ABU53070
ID ABU53070 standard; protein; 951 AA.
AC ABU53070;
XX
XX 15-APR-2003 (first entry)
XX
XX Intracellular trafficking-associated DKFZphtes3_lg13 homologue #3.
XX Human; gene therapy; vaccine; disease treatment; detection.
XX Homo sapiens.
XX WO200112659-A2.
XX
XX 22-FEB-2001.
XX
XX 18-AUG-2000; 2000WO-IB001496.
XX
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX WPI; 2001-327840/34.
XX
XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX
XX Example III; Page 667; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention

XX Sequence 951 AA;
SQ
Query Match 7.3%; Score 110; DB 4; Length 951;
Best Local Similarity 20.6%; Pred. No. 5.3;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;
QY 7 EQTVVVKSAIETADGALDLYNKYLDQVWPWKTFDETIKELSRFKQY-----SOEA 58
DB 394 EKQMEKVKQAKEMQETLK--KKLLDOEAKLKELENTALELSQEKQFNAKMLEMAQAN 451
QY 59 SVLVGDIKVLMDSQDKYFEATQTVY--EWCQGVVT-----QLLSAYILLFDEYN 105
DB 452 SAGISDAVSRLETNQKEIESLTVHRRELNDVISWEKLNQQAELQEIHIQLOEKE 511
QY 106 EKKASQAQKIDILI-----RILDDGVKK---LNEAQKSLITSSQSFNNAKSL 149
DB 512 QEVAELKQKILLFGCEKEEMNKETWLKSEGVQDQTTLNELOQLKQKSAHVNSLAQDET 571
QY 150 ALDSQL-----TNDFSEKSSYFQSQVDRIKEAYAGAAAGIVAGPFGILIIYSIAAGVI 203
DB 572 KKAHLEKLEVDLNLKSKENTFLQEQQLVELKMLAE----- 607
QY 204 EGKLIPELNRLKTVQNFFTSLSATVKQANKOIDAAKL---KLATEIAAIGEI---KTET 257
DB 608 DRKXVSELTSLKTTDEEFQSLKSSHEKSNKSLSEDKSLEFKLSELAIQIDICCKTE- 666
QY 258 ETTRFYVDYDDLMLSLKGAAKKMIN 283
DB 667 -----ALLEAKTNELIN 678

RESULT 17
ABU53071
ID ABU53071 standard; protein; 953 AA.
AC ABU53071;
XX
XX 15-APR-2003 (first entry)
XX
XX Intracellular trafficking-associated DKFZphtes3_lg13 homologue #4.
XX Human; gene therapy; vaccine; disease treatment; detection.
XX Homo sapiens.
XX WO200112659-A2.
XX
XX 22-FEB-2001.
XX
XX 18-AUG-2000; 2000WO-IB001496.
XX
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX WPI; 2001-327840/34.
XX
XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX
XX Example III; Page 668; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of

CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 953 AA;

Query Match 7.3%; Score 110; DB 4; Length 953;
Best Local Similarity 20.6%; Pred. No. 5.3;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;

Qy 7 EOTVEVVKSAIETADGALDLYNKLDQVWPW-KTFDETIKELSRFKQBY-----SQEA 58
Db 484 EKMEKVKQAKEMQETLK--KLLDQEAKLKKELENTALELSQKEQFNKMLEMAQAN 541
Qy 59 SVLVGDIKVLMDSDQKYFEATQVY--EMCGVVT-----QLLSAYILLFDEYN 105
Db 542 SAGISDAVSRLTQKEQIESLTVHRRELNDVISWEKKLNQQAELQEIHIQLOEKE 601
Qy 106 EKASAKQDILI-----RILDDGVKK--LNEAQSKLLTSSQSFNNASGKLL 149
Db 602 QEVAEKQKILLFGCEKEEMNKETWLKEGVKQDITLNEQLQEKQSAHVNSLAQDET 661
Qy 150 ALDSQL-----TNDFSEKSSYFQSDVRIRKEAYAGAAAGIVAGPPGLIISYIAAGVI 203
Db 662 KKAHLEKLEVDLNLKSLKENTFLOQLVELKMLAEE----- 697
Qy 204 EGKLIPELNNRLKTVQNFTSLSATVQKANKIDAAKL---KLATEIAAIGEI---KTET 257
Db 698 DKRKVSELTSLKTTDEEFQSLKSSHEKSNKSLDKSLEPKKJSEELAIQLDICKCKTE- 756
Qy 258 ETRFYVDYDMLSLKGAAKMIN 283
Db 757 -----ALLEAKTNELIN 768

RESULT 18
ABUS3077
ID ABUS3077 standard; protein; 961 AA.
XX
AC ABUS3077;
XX
DT 15-APR-2003 (first entry)
XX
DE Intracellular trafficking-associated DKFzptes3_1g13 homologue #10.
XX
KW Human; gene therapy; vaccine; disease treatment; detection.
XX
OS Homo sapiens.
XX
PN WO200112659-A2.
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-IB001496.
XX
PR 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
PA
XX
XX Wiemann S;
XX
XX WPI; 2001-327840/34.
XX
XX Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.
PT
XX
XX Example III; Page 673; 1095pp; English.
PS
XX
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC

CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 961 AA;

Query Match 7.3%; Score 110; DB 4; Length 961;
Best Local Similarity 20.6%; Pred. No. 5.3;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;

Qy 7 EOTVEVVKSAIETADGALDLYNKLDQVWPW-KTFDETIKELSRFKQBY-----SQEA 58
Db 297 EKMEKVKQAKEMQETLK--KLLDQEAKLKKELENTALELSQKEQFNKMLEMAQAN 354
Qy 59 SVLVGDIKVLMDSDQKYFEATQVY--EMCGVVT-----QLLSAYILLFDEYN 105
Db 355 SAGISDAVSRLTQKEQIESLTVHRRELNDVISWEKKLNQQAELQEIHIQLOEKE 414
Qy 106 EKASAKQDILI-----RILDDGVKK--LNEAQSKLLTSSQSFNNASGKLL 149
Db 415 QEVAEKQKILLFGCEKEEMNKETWLKEGVKQDITLNEQLQEKQSAHVNSLAQDET 474
Qy 150 ALDSQL-----TNDFSEKSSYFQSDVRIRKEAYAGAAAGIVAGPPGLIISYIAAGVI 203
Db 475 KKAHLEKLEVDLNLKSLKENTFLOQLVELKMLAEE----- 510
Qy 204 EGKLIPELNNRLKTVQNFTSLSATVQKANKIDAAKL---KLATEIAAIGEI---KTET 257
Db 511 DKRKVSELTSLKTTDEEFQSLKSSHEKSNKSLDKSLEPKKJSEELAIQLDICKCKTE- 569
Qy 258 ETRFYVDYDMLSLKGAAKMIN 283
Db 570 -----ALLEAKTNELIN 581

RESULT 19
ABR61599
ID ABR61599 standard; protein; 2228 AA.
XX
AC ABR61599;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human golgin-245 splice variant 1 polypeptide.
XX
KW Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.
XX
OS Homo sapiens.
XX
PN WO2003087403-A2.
XX
PD 23-OCT-2003.
XX
PF 16-APR-2003; 2003WO-EF003958.
XX
PR 16-APR-2002; 2002EP-00008553.
PR 16-APR-2002; 2002US-0372424P.
XX
XX (EVOT-) EVOTEC NEUROSCIENCES GMBH.
XX
XX Von Der Kammer H, Pohlner J;
PI WPI; 2003-845345/78.
XX N-PSDB; ACF58150.
XX
XX Diagnosing or prognosticating a neurodegenerative disease comprises
PT comparing a level or activity of a transcription or translation product
PT of golgin-245 gene in a test sample with a reference value of a known
PT disease or health status.
PT


```
XX PR Claim 18; Fig 9; Opp; English.
XX PA
XX PI The invention relates to diagnosing or prognosticating a
XX CC neurodegenerative disease in a subject, or determining whether a subject
XX CC is at increased risk of developing the disease. The method involves
XX CC determining a level and/or an activity of a transcription or translation
XX CC product of a gene coding for golgin-245 and/or a fragment, derivative or
XX CC variant in a sample from the subject and comparing the level and/or the
XX CC activity to a reference value. The composition and methods are useful in
XX CC diagnosing, preventing and/or treating neurodegenerative diseases, such
XX CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
XX CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
XX CC useful for screening, testing and validating compounds, agents and
XX CC modulators in the development of diagnostics and therapeutics to treat
XX CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
XX CC antibodies are used for detecting the pathological state of a cell in a
XX CC sample from a subject. The present sequence represents a human golgin-245
XX CC splice variant 3 polypeptide
XX SQ Sequence 2250 AA;
Query Match 7.3%; Score 110; DB 7; Length 2250;
Best Local Similarity 20.6%; Pred. No. 16;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;
QY 7 EQTVVVVKSATETADGALDLYNKYLDQVTPW-KTFDETIKELSRFKQY-----SQEA 58
Db 975 EKQMEVKQKAKEMQETLK--KLLDQEAALKKELENTALELSQKQFNKMLEMAQAN 1032
QY 59 SVLVGDIKVLMSODKYFEATQTVY--EWCQVVT-----QLLSAYILLFDEYN 105
Db 1033 SAGISDAVSRLETNQEQIESLTVHRRLNDVSIWEKLNQQAELQBIHEIQLQKE 1092
QY 106 EKASAKQDILI-----RILDDGVKK---LNEAQSKLLTSSQSFNNASCKLL 149
Db 1093 QEVAELKQKILLFCCEKEENKKEITWLKEGVKQDITLNEQLQKQSAHVNSLAQDET 1152
QY 150 ALDSQL-----TNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPPGLIISYIAAGVI 203
Db 1153 KLKAHLEKLEVDLNLKSLKENTFLQEQVLVELKMLAE-----1188
QY 204 EGKLIPELNNRLKTVQNFFTSLSATVKQANKDIAAKL---KLATEIAAIGEI---KTET 257
Db 1189 DKRKVSELTSLKLTDEEFQSLKSSHEKSNKSLKEDKSLFEPKGLSELAIQDICKCKTE- 1247
QY 258 ETTRFYVDYDDMLSLKGAACKWIN 283
Db 1248 -----ALLEAKTNELIN 1259
RESULT 23
ABR61602
ID ABR61602 standard; protein; 2252 AA.
XX AC ABR61602;
XX DT 15-JAN-2004 (first entry)
XX DE Human golgin-245 splice variant 4 polypeptide.
XX KW Golgin-245; neurodegenerative disease; neuroprotective; neurotropic;
XX OS antiparkinsonian; anticonvulsant; gene therapy; human; variant.
XX PN Homo sapiens.
XX PD W02003087403-A2.
XX PF 23-OCT-2003.
XX PR 16-APR-2003; 2003WO-EP003958.
XX PR 16-APR-2002; 2002EP-00008553.
```

```
PR 16-APR-2002; 2002US-0372424P.
XX PA (EVOT-) EVOTEC NEUROSCIENCES GMBH.
XX PI Von Der Kammer H, Pohlner J;
XX CC WPI; 2003-845345/78.
XX CC N-PSDB; ACF58153.
XX PT Diagnosing or prognosticating a neurodegenerative disease comprises
XX CC comparing a level or activity of a transcription or translation product
XX CC of golgin-245 gene in a test sample with a reference value of a known
XX CC disease or health status.
XX PS Claim 18; Fig 11; Opp; English.
XX CC The invention relates to diagnosing or prognosticating a
XX CC neurodegenerative disease in a subject, or determining whether a subject
XX CC is at increased risk of developing the disease. The method involves
XX CC determining a level and/or an activity of a transcription or translation
XX CC product of a gene coding for golgin-245 and/or a fragment, derivative or
XX CC variant in a sample from the subject and comparing the level and/or the
XX CC activity to a reference value. The composition and methods are useful in
XX CC diagnosing, preventing and/or treating neurodegenerative diseases, such
XX CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
XX CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
XX CC useful for screening, testing and validating compounds, agents and
XX CC modulators in the development of diagnostics and therapeutics to treat
XX CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
XX CC antibodies are used for detecting the pathological state of a cell in a
XX CC sample from a subject. The present sequence represents a human golgin-245
XX CC splice variant 4 polypeptide
XX SQ Sequence 2252 AA;
Query Match 7.3%; Score 110; DB 7; Length 2252;
Best Local Similarity 20.6%; Pred. No. 16;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;
QY 7 EQTVVVVKSATETADGALDLYNKYLDQVTPW-KTFDETIKELSRFKQY-----SQEA 58
Db 975 EKQMEVKQKAKEMQETLK--KLLDQEAALKKELENTALELSQKQFNKMLEMAQAN 1032
QY 59 SVLVGDIKVLMSODKYFEATQTVY--EWCQVVT-----QLLSAYILLFDEYN 105
Db 1033 SAGISDAVSRLETNQEQIESLTVHRRLNDVSIWEKLNQQAELQBIHEIQLQKE 1092
QY 106 EKASAKQDILI-----RILDDGVKK---LNEAQSKLLTSSQSFNNASCKLL 149
Db 1093 QEVAELKQKILLFCCEKEENKKEITWLKEGVKQDITLNEQLQKQSAHVNSLAQDET 1152
QY 150 ALDSQL-----TNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPPGLIISYIAAGVI 203
Db 1153 KLKAHLEKLEVDLNLKSLKENTFLQEQVLVELKMLAE-----1188
QY 204 EGKLIPELNNRLKTVQNFFTSLSATVKQANKDIAAKL---KLATEIAAIGEI---KTET 257
Db 1189 DKRKVSELTSLKLTDEEFQSLKSSHEKSNKSLKEDKSLFEPKGLSELAIQDICKCKTE- 1247
QY 258 ETTRFYVDYDDMLSLKGAACKWIN 283
Db 1248 -----ALLEAKTNELIN 1259
RESULT 24
ABU25599
ID ABU25599 standard; protein; 742 AA.
XX AC ABU25599;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #11126.
```


Db 425 VTETFRSISGLVEMANDLEKRALAQVSKDVEAL--NQVNEAIQOVSEIARQOQETI 482
Qy 59 SVLVGDIKVLMLDSQDKYFEATQTVVEMCGVVTQLLSAYILLDFEYNEKASQKDI--- 115
Db 483 NEITDGMRLVAQTSE-----ESVRAMEFSGAVTEVVS-----IANEGSQKQDEALKRIEDI 534
Qy 116 --LIRILDDGVKVLNEAQKSL-----LTSSQSPNNASGKLLALDS----- 153
Db 535 QHMSRIETVSKVAEMSRNIEITNVITSIASQTN-----LLALNAAIEAARAGEAGRGF 590
Qy 154 -----QUTNDFSEKSSYFQSOVDRIR---KEAVAGAAAGI-VAGPFGLLIISYSTAAG 201
Db 591 AVVAQEIRKLAESKQAADNIKSIIDKITDEIKEAVEATKEGVSIGESSETLRDTIGYL 650
Qy 202 VIGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAUKLATEIAAIGETITETTR 261
Db 651 ANIATLIQETSERMTTVKE-----QIVRTQBEVDKALRALENLAASABETTASAEVS 703
Qy 262 FYVDYDDLMLSLKGAKKMINTCNEYQORHGK 294
Db 704 SAIEQQTAAIEELRRAAQELKDMVGRMRQIVGK 736

RESULT 26
ABU53079
ID ABU53079 standard; protein; 929 AA.
AC ABU53079;
XX
XX
DT 15-APR-2003 (first entry)
XX
DE Intracellular trafficking-associated DKFZphtes3_1913 homologue #12.
XX Human; gene therapy; vaccine; disease treatment; detection.
XX Homo sapiens.
XX
XX W0200112659-A2.
PN
XX
XX 22-FEB-2001.
XX
XX 18-AUG-2000; 2000WO-IB001496.
XX
XX 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX
XX WPI; 2001-327840/34.
XX

XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX
XX Example III; Page 674-675; 1095pp; English.
XX
XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention
XX
XX Sequence 929 AA;
XX
XX Query Match 7.28; Score 109; DB 4; Length 929;
XX Best Local Similarity 20.6%; Pred. No. 6.2;

Matches 67; Conservative 59; Mismatches 106; Indels 94; Gaps 12;
Qy 7 EQTVEVVKAAIETADGALDLYNKYLDQVTPW-KTFDETIKELSRPKQEY-----SOBA 58
Db 1 EKVKQKAKEMOET-----LKKLLDQEAQKLKELENTALELSQEKQFNAKMLEMAQAN 54
Qy 59 SVLVGDIKVLMLDSQDKYFEATQTVY--EWCQGVVT-----QLLSAYILLDFEYN 105
Db 55 SAGISDAVSRLETNQKEQIESLTVHRRELNDVISIWEKKLNQQAEBELQIHLQLOEKE 114
Qy 106 EKKAQAQKDIIL-----RILDDGVKK-----LNEAQSLLTSSQSPNNASGKLL 149
Db 115 QEVAELKQKILLFGCEKEEMNKETITWKEBQVQDITLNELOQLKQKSAHVNSLAQDET 174
Qy 150 ALDSQL-----TNDSEKSSYFQSOVDRIRKEAYAGAAAGIVAGPFGLLIISIAAGVI 203
Db 175 KLKAHLKLEKLEVDLNLKSLKENTFLQEQVLKMLAEE----- 210
Qy 204 EGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAUKL-----KLATETIAAIGETI---KTET 257
Db 211 DRKVSLETSKLTDTDEBFQSLKSSHEKSNKSLSEKSLSEKSLSELAIQDIDICCKYTE- 269
Qy 258 ETTRFYVDYDDLMLSLKGAKKMIN 283
Db 270 -----ALLEAKTNELIN 281

RESULT 27
ABB59245
ID ABB59245 standard; protein; 1048 AA.
XX
XX ABB59245;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4527.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX W0200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL03348.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.

XX Disclosure; SEQ ID NO 4527; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1048 AA;

Query Match 7.2%; Score 108.5; DB 4; Length 1048;
Best Local Similarity 21.6%; Pred. No. 7.9; Indels 53; Gaps 12;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;
QY 24 LDLYNKYLDQVIPWKTDETIKLSRFKQYSEASVLDGDIKVLMLSDQDKYFEATQTV 83
DB 159 LELNNEELCDLL-----STDDTK--IRIFDDSTKSGVILQGLEIPVHSDVDVYKLEKG 213
QY 84 YEWGCVVTLQSLAYILLFDEYNEKASAKDILIRILDDGVK-----KLENAQKSLTSS 138
DB 214 KERRKTATTLNNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNVLVDLAGS 265
QY 139 QSFNNASK-----LLALDSOLTN--DFSEKSSYFQSOVDRIKRAYAGAA 182
DB 266 ENVKAGNEKIRVRETWNINQSLTLGRVITLVDRAPHVPYRESKLTLLQESLGRT 325
QY 183 -----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNRL--KTVQNFPTLSATVK 230
DB 326 KTSIIATISPGHKDIEETLSLEYAHRANKIQNK--PEVNQKLTKTVLKEYTE---EID 380
QY 231 QANKDIDAAKLKLATEIA--AIGEIKTETETTRFVVDYDDLMLSLK 275
DB 381 KLRDLMAARDKNGIYLAETTYGEITLKESQNLRELNEKMLLLKALK 427

RESULT 28

AAG67418
ID AAG67418 standard; protein; 1066 AA.
XX
AC AAG67418;
DT 13-NOV-2001 (first entry)
XX
DE Amino acid sequence of bimC homologue, KLP61.
XX
KW bimC; kinesin related protein; fungal viability; antifungal; KLP61;
KW fungal infection.
XX
OS Drosophila melanogaster.
XX
FN US6284480-B1.
XX
PD 04-SEP-2001.
XX
PF 03-APR-2000; 2000US-00541782.
XX
PR 03-APR-2000; 2000US-00541782.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
XX Nislow CE, Sakowicz R, Beraud C;
XX
XX WPI: 2001-540724/60.
XX
DR N-PSDB; AAH78012.
XX
XX Identifying a modulator, e.g. antifungal agent, of a target protein
XX comprising bimC or its fragment by determining enzymatic activity of a
XX reaction, in the presence and absence of the compound, that uses ADP or
XX phosphate produced by bimC.
XX
XX Disclosure; Fig 8; 47pp; English.

XX The present sequence represents a bimC homologue, designated KLP61. BimC
XX is a kinesin related protein, which is essential for fungal viability.
XX The specification describes a method of identifying modulators of bimC.
XX The method comprises adding a test agent to a mixture comprising bimC
XX protein that directly or indirectly produces ADP or phosphate, subjecting
XX the mixture to an enzymatic reaction that uses the ADP or phosphate, and
XX determining the enzymatic activity in presence and absence of test

CC compound. A change in the activity level between the presence and absence
CC of the candidate agent indicates a modulator of the target protein
CC function. The method is useful for identifying a modulator, e.g.
CC antifungal agents, of bimC. The modulators can be used, for example, to
CC inhibit the growth or spread of fungi, mould, fruit flies, etc. . The
CC modulators can be used for preventing and treating infections caused by
CC Chytridiomycetes, Hyphochytridiomycetes, Basidiomycetes, and
CC Oomycetes, Zygomycetes, Ascomycetes, and Basidiomycetes

XX Sequence 1066 AA;

Query Match 7.2%; Score 108.5; DB 4; Length 1066;
Best Local Similarity 21.6%; Pred. No. 8.1; Indels 53; Gaps 12;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;
QY 24 LDLYNKYLDQVIPWKTDETIKLSRFKQYSEASVLDGDIKVLMLSDQDKYFEATQTV 83
DB 159 LELNNEELCDLL-----STDDTK--IRIFDDSTKSGVILQGLEIPVHSDVDVYKLEKG 213
QY 84 YEWGCVVTLQSLAYILLFDEYNEKASAKDILIRILDDGVK-----KLENAQKSLTSS 138
DB 214 KERRKTATTLNNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNVLVDLAGS 265
QY 139 QSFNNASK-----LLALDSOLTN--DFSEKSSYFQSOVDRIKRAYAGAA 182
DB 266 ENVKAGNEKIRVRETWNINQSLTLGRVITLVDRAPHVPYRESKLTLLQESLGRT 325
QY 183 -----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNRL--KTVQNFPTLSATVK 230
DB 326 KTSIIATISPGHKDIEETLSLEYAHRANKIQNK--PEVNQKLTKTVLKEYTE---EID 380
QY 231 QANKDIDAAKLKLATEIA--AIGEIKTETETTRFVVDYDDLMLSLK 275
DB 381 KLRDLMAARDKNGIYLAETTYGEITLKESQNLRELNEKMLLLKALK 427

RESULT 29

ABG72693
ID ABG72693 standard; protein; 1066 AA.
XX
AC ABG72693;
DT 06-MAR-2003 (first entry)
XX
DE Fruitfly KLP61, a homologue of BimC, a kinesin related protein.
XX
KW Fruitfly; bimC; kinesin related protein; mitosis; antifungal;
KW Candida infection; fusaria infection; zygomycosis; cryptococcosis;
KW histoplasmosis; KLP61.
XX
OS Drosophila melanogaster.
XX
FN US6468760-B1.
XX
PD 22-OCT-2002.
XX
PF 28-NOV-2000; 2000US-00723820.
XX
PR 03-APR-2000; 2000US-00541782.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
XX Nislow CE, Sakowicz R, Beraud C;
XX
XX WPI: 2003-147102/14.
XX
DR N-PSDB; ABX14663.
XX
XX Novel kit for screening for modulators of bimC, comprises a biologically
XX active bimC, its fragment or homolog, and instructions for testing bimC
XX activity.
XX
XX Disclosure; Fig 8; 47pp; English.

CC The invention relates to a kit for screening for modulators of b1mC,
CC comprises a biologically active b1mC (a kinesin related protein essential
CC for mitosis), its fragment or homologue, and instructions for testing
CC b1mC activity. The kit is used for screening for modulators of b1mC. The
CC kit is also useful for identifying therapeutic agents (antifungal agents
CC against Candida infections, fusaria infections, zygomycosis,
CC cryptococcosis and histoplasmosis) and agricultural compounds. The
CC present sequence represents the fruitfly b1mC homologue, KUP61
XX
SQ Sequence 1066 AA;
SQ

Query Match	7.2%	Score 108.5;	DB 6;	Length 1066;
Best Local Similarity	21.6%;	Pred. No. 8.1;		
Matches	62;	Conservative 62;	Mismatches 110;	Indels 53; Gaps 12;

Qy	24	LDLNNKYLDQVWPWKTDEITIKELSRPKQESQESASVLVGDIKVLLMDSQDKYFEATQTV	83
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	159	LELYNEELCDLL---STDDTTK--IRIFDSTKKGSVIIQGLEEIPVHSKDDVYKLLERG	213
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	84	YEWCGVVTQLLSAYILLFDEYNEKKASAKQDILIRLLDDGVK-----KLANEAKSLLTSS	138
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	214	KERRKTATYLLMNA-----QSSRSHTVFSIVHIRENGIEGDMKIGKLNULDLAGS	265
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	139	QSFNNASGK-----LLALDLSQLTN--DFSEKSSYFQSQVDRIKEAYAGAA	182
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	266	ENVSKAGNEKGIRVRETWNINQSLTLGRVITALVDRAHPVPRESKLTILLQESLGRT	325
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	183	-----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNNRL--KTVQNFFTSLSATVK	230
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	326	KTSIIATISPGHKDIEETLSLEYAHRAKNIQNK--FEVNQKLTKTIVLKEYTE--EID	380
		: : : : : : : : : : : : : : : : : : : : : : : :	
Qy	231	QANKDDAACKLKLATEIA--AIGBIKTETETTRFYVDYDDMLMSLLK	275
		: : : : : : : : : : : : : : : : : : : : : : : :	
Db	381	KLKRDLMARDKXGIVLAEEFTYGBITLKLSONRELNEKMLLKALK	427
		: : : : : : : : : : : : : : : : : : : : : : : :	

RESULT 30
ADG98862
ID ADG98862 standard; protein: 1066 AA.

11-MAR-2004 (first entry)

DE Fruit fly KLP61 polypeptide.

KW Adenosine diphosphate; ADP; phosphate; bimC; kinesin related protein;
KW KRP; fruit fly; KLP61.

XX Drosophila melanogaster.

PN US6627408-B1.

30-SEP-2003.

PF 11-OCT-2002; 2002US-00270085.

PR 03-APR-2000; 2000US-00541782.

PR 28-NOV-2000; 2000US-00723820.

PA (CYTO-) CYTOKINETICS INC.

PI Nislow CE, Sakowicz R, Beraud C;

DR WPI; 2003-895513/82.

DR N-PSDB; ADG98861.

Identification of candidate agent as modulators of tar protein function by adding candidate agent to mixture of target protein, subjecting mixture to enzymatic reaction, and determining level of activity of PT enzymatic reaction

PS Disclosure; SEQ ID NO 8; 47pp; English.

The invention relates to a candidate agent that binds to a target protein. Identification of a candidate agent as a modulator of function of a target protein involves adding the candidate agent to a mixture comprising the target protein that directly or indirectly produces adenosine diphosphate (ADP) or phosphate under conditions which normally allow the production of ADP or phosphate, subjecting the mixture to an enzymatic reaction that uses the ADP or phosphate as a substrate under conditions which normally allow the ADP or phosphate to be utilised, and determining the level of activity of the enzymatic reaction where a change in the level between the presence and absence of the candidate agent indicates modulation of the target protein function. The method is useful for identifying a candidate agent as a modulator of function of a target protein. The method provides a high throughput system for detecting modulators of bImc enzyme activity. This sequence represents the fruit fly KUP61 polypeptide, which is a homologue of the kinein related protein (KRP) bImc, a target protein of the invention.

Query Match	7.2%;	Score 108.5;	DB 7;	Length 1066;
Best Local Similarity	21.6%;	Pred. No. 8.1;		
Matches 62;	Conservative 62;	Mismatches 110;	Indels 53;	Gaps 12;
Qy	24	LDLYNKYLDQVWPWKTDEIIRKLSRRKQESQBSASVLVGDIDIKVLLMDSQDKYFEATQTV	83	
Db	159	LELYNEELCDLL---STDDTTK---IRFDSTTKGSGVIOGLEBIPVHSKDDVVKLLFKG	213	
Qy	84	YEMCGVVQTLLSAVILAFDEYNEKKAQAQDKILIRILDGKVK----	138	KLNEAQSLLTSS
Db	214	KERRKTATTLNNA-----QSSRSHTVFSIVVHIRENGIEGDMUKIGKINLVDL	265	LAGS
Qy	139	QSFNNASGK-----LLALDSOLTN--DFSEKSSYPQSOVDRIKRAYAGAA	182	
Db	266	ENWSKAGNEKGI RVETWNIQSLLTLCRVITALVDRAHPVPYRESKLTLLQESLGR	325	T
Qy	183	-----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNNRL--KTQVNFETSL	230	ATVK
Db	326	KTSIIATISPGHKDIBETSLTSLYAHRAKNIQNK--PEVQKLTKTIVLKEYTE---	380	EID
Qy	231	QANKDIDAAKLKLAETEIA--AIGBIKTETETTRFYVDYDDLMLSLK	275	
Db	381	KLKRDLMARQKNGIYLAEEYTGSIITUKLSONRELNEKMLLKALK	427	

RESULT 31

ABU43311

ID ABU43311 standard; protein; 3692 AA.

AC ABU43311;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #28838.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Staphylococcus epidermidis.

PN WO200277183-A2.

PD 03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001: 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX
02 /E/ T T -) E/ T T P A P H A B M T N C
XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA47181.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 71235; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 3692 AA;

Query Match 7.2%; Score 108.5; DB 6; Length 3692;
Best Local Similarity 21.5%; Pred. No. 40;
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;

QY 11 EVVKSATETADGALDLYNKYLQVVPKTFDETIKLSR-----FKQESQ-EA 58
DB 564 EQNKDIIPSNYTLASINKY-----NKLKRAQTVDLTETNTFPNQRYSQTOI 612
QY 59 SVLVGDIKVLMM-----DSQKYPEATQTVYECGVTQLLSAYILLDFEYNEKKASA 111
DB 613 DDLHLQLTTLINRVASREINDKAQEMTDAVYDSTELTTEKDT---LVDQIENHKNEI 669
QY 112 QKDILRLDDGVKLNKAEAKSLTSS-----QSFNNASGKLALDLSQLTNDFS 160
DB 670 SNNDDELTDGVERVKEAGLHTLES DTPHVTKPNARQVNNRA-----DQCKTLIRN 723
QY 161 EKSSYFOSQVDRIRK-EAYAGAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTQV 219
DB 724 NHEATTEQNEAIRQVEAHSSDAIAKIGE-----AETDTTVNEAR 763
QY 220 NFFTSLSAT-VKQANKIDIAKLAETAAIGETETETTFYVDYDMLSLKGA 278
DB 764 DNGTKLIATDVPNPTKAEAE---RAAVTNSANSKIDINNTOATLDERNDALVNRSK 820
QY 279 KQMTNCTNEQY-----QRHGKTKLFEVP 301
DB 821 DEALQNTAQGNDVDVTEAQNGNTNTIQVP 851

RESULT 32
ABP40235
ID ABP40235 standard; protein; 3696 AA.
XX
XX AC ABP40235;
XX 24-JUL-2002 (first entry)
DT
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.
DE
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
OS
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-00134001.
XX
XX 14-AUG-1997; 97US-0055779P.
PR 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
PI
XX WPI; 2002-381255/41.
DR N-PSDB; ABN92780.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
XX Disclosure; SEQ ID NO 5080; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly *S. epidermidis* infections. The sequences can be used to
CC screen for compounds able to interfere with the *S. epidermidis* life cycle
CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
XX SQ Sequence 3696 AA;

Query Match 7.2%; Score 108.5; DB 5; Length 3696;
Best Local Similarity 21.5%; Pred. No. 40;
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;

QY 11 EVVKSATETADGALDLYNKYLQVVPKTFDETIKLSR-----FKQESQ-EA 58
DB 568 EQNKDIIPSNYTLASINKY-----NKLKRAQTVDLTETNTFPNQRYSQTOI 616
QY 59 SVLVGDIKVLMM-----DSQKYPEATQTVYECGVTQLLSAYILLDFEYNEKKASA 111
DB 617 DDLHLQLTTLINRVASREINDKAQEMTDAVYDSTELTTEKDT---LVDQIENHKNEI 673
QY 112 QKDILRLDDGVKLNKAEAKSLTSS-----QSFNNASGKLALDLSQLTNDFS 160
DB 674 SNNDDELTDGVERVKEAGLHTLES DTPHVTKPNARQVNNRA-----DQCKTLIRN 727
QY 161 EKSSYFOSQVDRIRK-EAYAGAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTQV 219
DB 728 NHEATTEQNEAIRQVEAHSSDAIAKIGE-----AETDTTVNEAR 767
QY 220 NFFTSLSAT-VKQANKIDIAKLAETAAIGETETETTFYVDYDMLSLKGA 278

Db 768 DNGTKLIATDPNPTKAEAA---RAAVTNSANSKIKDINNNTQATLDERNDATAVNRSK 824

Qy 279 KMINTCNEYQ-----QRHGKKTLEFVP 301

Db 825 DEAIQININTAGNDVDVTEAQNNGTNTIQQVP 855

RESULT 33

ABBB77986

ID ABB77986 standard; protein; 1292 AA.

AC ABB77986;

XX

XX

DT 22-OCT-2002 (first entry)

XX

DE Amino acid sequence of an Arabidopsis RAD50 homologue.

XX

XX Nucleic acid integration; homologous recombination; telomeric region;

KW RAD50.

XX

XX Arabidopsis thaliana.

OS

XX

PN EPI1217074-A1.

XX

PD 26-JUN-2002.

XX

XX 22-DEC-2000; 2000EP-00204693.

PF

XX

PR 22-DEC-2000; 2000EP-00204693.

XX

XX (UYLE-) RIJKSUNIV LEIDEN.

PA (BINA-) STICHTING BINAIR VECTOR SYSTEM.

XX

PI Hooykaas PJJ, Van Attikum H, Bundock P;

XX

XX WPI; 2002-550405/59.

DR

XX

PT Directing integration of nucleic acid of interest to a sub-telomeric region in an eukaryote with preference for non-homologous recombination, by steering an integration pathway towards homologous recombination.

XX

PS Disclosure; Fig 5; 63pp; English.

XX

CC The specification describes a method for directing integration of a nucleic acid of interest to a pre-determined site, where the nucleic acid has homology at or around the pre-determined site, in a eukaryote with a preference for non-homologous recombination. The method comprises steering an integration pathway towards homologous recombination. The method is useful for directing integration of a nucleic acid of interest to a subtelomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination. The nucleic acid of interest comprises an inactive gene to replace an active gene, or vice versa, is a portion of a gene delivery vehicle, confers a desired property to the eukaryotic cell, or encodes a therapeutic proteinaceous substance or a substance conferring resistance for an antibiotic substance to a cell. The method is useful for improving gene targeting efficiency. The method is useful in the replacement of an active gene by an inactive gene, for e.g. for the inactivation of genes controlling undesired side branches of metabolic pathways, to increase the quality of bulk products such as starch, or to increase the production of specific secondary metabolites or to inhibit formation of unwanted metabolites, and also to inactivate genes controlling senescence in fruits and flowers or that determine flower pigments. The method is also useful for replacing an inactive gene by an active gene. For e.g. the replacement of a defective p53 by an intact p53. Many tumours acquire a mutation in p53 during their development which renders it inactive and often correlates with a poor response to cancer therapy. By replacing the defect p53 by an intact p53, e.g. through gene therapy, conventional anti cancer therapy have better changes of succeeding. The method is also useful for therapeutic proteinaceous substance integration. A tumoricidal gene can be delivered to a pre-determined site present only in e.g. proliferating cells, or present only in tumour cells, e.g. to the site where a tumour antigen is expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved

CC in non-homologous recombination

XX

SQ Sequence 1292 AA;

Query Match 7.1%; Score 107.5; DB 5; Length 1292;

Best Local Similarity 17.2%; Pred. No. 13;

Matches 69; Conservative 77; Mismatches 138; Indels 117; Gaps 13;

Qy 2 TSIFAQTVVEVKSJAIETADGALDLYNKYLDQVVPKWT-PDE----- 42

Db 281 STLFEKQQRQYAAALPEENED-----TTEELKEWKSKEERLALGTKIRKMEREMVD 332

Qy 43 ---TIKELSRFKQESQESVL---VGDIKVLAMDSDQKYFEATQTVYEWGVVTVQLLSAY 97

Db 333 TETTISLHNKNTNYMLEISKLQTEAEAHMLLNKRDSTIQTFFHNLGNVSTPPSTE 392

Qy 98 ILIFDEYNEKKSAAQKIDILIRILDGKVLKNEAKSLLTSSQSFNNASGKLALDSQ--- 154

Db 393 VVL--NLTWRIKSRGLGELEMDLLDK--KKSNET--ALSTAWDCYMDANDRWKSIEAQKBA 446

Qy 155 -----LTNDFSEKSSY-----FQSQ 169

Db 447 KDEIKMGISKRIEKEIERDSFEFEISTVDVKQTDEREKQVQVELEKTKQNSRGEFSK 506

Qy 170 VDIRKEAYA-----GAAAGIVAGPFGLLIYSYI-----AAGVIEGKLIPELNNR- 214

Db 507 IEQKHIEIVSLEHKIKTLNRERDVMAGDAEDRLLTRIDECKDIRGVGLKRLPEKDKMR 566

Qy 215 -----LKTQNFPTSLSATVKQANKDIDAALKLATEIAAIGEIKTETTRFVVDYD-- 267

Db 567 EIVQALRSIEREYDDLKLSREAKEVNMLOMKIQEVNNSLFFHNKDTESKRYIESKLQ 626

Qy 268 -----DLMLSLKGAAGKMMTCNEYQQRHGKKTLE 299

Db 627 ALKQESVTIDAYPKLLESACKDRDKREYNWANGMRQMEF 667

RESULT 34

ABBS3560

ID ABR53560 standard; protein; 1875 AA.

XX

AC ABR53560;

XX

DT 20-JUN-2003 (first entry)

XX

DE Protein sequence #SEQ ID 1985.

XX

KW Multiprotein complex; eukaryote; drug target; diagnosis.

XX

OS Saccharomycetes cerevisiae.

XX

PN EP1258494-A1.

XX

PD 20-NOV-2002.

XX

PF 20-DEC-2001; 2001EP-00130253.

XX

PR 15-MAY-2001; 2001EP-00111774.

XX

XX (CELL-) CELLZOME AG.

XX

PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

PI Marzioch M, Schultz JD, Superti-Furga GD;

XX

XX WPI; 2003-250078/25.

DR N-PSDB; ACC61602.

XX

XX New isolated protein complexes useful for diagnosing a disease or disorder, or as a target for an active agent of a pharmaceutical, preferably a drug target in the treatment or prevention of disease or disorder.

XX

PS Disclosure; SEQ ID NO 1985; 17pp + Sequence Listing; English.

XX The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM
XX
SQ Sequence 1875 AA;
Query Match 7.1%; Score 107.5; DB 6; Length 1875;
Best Local Similarity 19.9%; Pred. No. 20;
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;
QY 7 EQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETI-----KELSRFKQEYSQAS 59
Db 966 EDKISLLKEQMFNLNLDLQKGMKEK--EKADFKKISILQNNKEVEAVKSEYKSL 1023
QY 60 VLVDGIKVLMSDQKYFEATQTVYEWGCVVTVLLSAYILLDFEYNEKASAKDI---- 115
Db 1024 KIQND-----LDQOTIYANTQNNYE-----QELQKHADVSKTISELR 1061
QY 116 -LIRILDGVKKLN-----EAQKSLTTSQSFNNAAGKLLA---LDSOLTNDPSEKSYFQ 167
Db 1062 EQHTYKGVQKVTNLNLRDQLENALKENKESWSQKESLLEQLDLSNRIEDLSQNKLLY 1121
QY 168 SQVDRIRKEAYAGAAGIVAGPFGLLISYSTAAGVIEGKLIPELNNRLKTVQNFFTLSA 227
Db 1122 DQI-----QIYTAADKEVNNSTNG-----PGLNNILITLRERDILT 1159
QY 228 TVQKANKDIDAAGKLAETAAIGETKTETTR-----FYVDYDDL-----LSLL 274
Db 1160 KTVVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNLL 1219
QY 275 KGAAKKMINTCNEYQQRHGKK 295
Db 1220 R---ESNITLRNELENNNKK 1237
RESULT 35
ADK64380
ID ADK64380 standard; protein; 1875 AA.
XX
AC ADK64380;
XX
DT 06-MAY-2004 (first entry)
XX
DE Disease treating protein complex-derived protein #1185.
XX
KW protein complex; drug target; diagnosis.
XX
OS Unidentified.
XX
XX EF1338608-A2.
XX
XX 27-AUG-2003.
XX
XX 20-DEC-2002; 2002EP-00102902.
XX
XX 20-DEC-2001; 2001EP-00130253.
XX
XX (CELL-) CELLZOME AG.
XX
XX Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
XX Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
XX Michon A, Leutwein C, Rick J;
XX

WPI; 2003-638460/61.
N-PSDB; ADK64381.

New proteins and protein complexes from eukaryotes, useful as targets in drug screening, or in diagnosing or screening for the presence of a disease or disorder, or a predisposition for developing a disease or disorder in a subject.

Disclosure; SEQ ID NO 2369; 13pp; English.

The invention relates to novel protein complexes comprising a first and a second protein, or its derivative, fragment, homologue or variant. The proteins are selected from given protein complexes, which are not defined in the specification. The variants are encoded by nucleic acids that hybridize to the nucleic acids encoding the proteins under low stringency conditions. The protein complexes are useful as targets for an active agent of a pharmaceutical. These protein complexes are particularly useful as drug targets for the treatment or preventing of a disease or disorder. The complexes and methods above are useful in diagnosing or screening for the presence of a disease or disorder. These are also useful in developing a drug for treatment or prevention of a disease or disorder. The molecule that modulates the amount, activity or protein components of the complex is useful for the manufacture of a medicament for the treatment or prevention of a disease or disorder. This sequence corresponds to a protein of the invention. (Note: the sequence data for this patent did not form part of the printed specification but was obtained from the EPO in electronic format).

Sequence 1875 AA;

Query Match 7.1%; Score 107.5; DB 7; Length 1875;
Best Local Similarity 19.9%; Pred. No. 20;
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;

QY 7 EQTVVVKSAIETADGALDLYNKYLDQVLPWKTFDETI-----KELSRFKQEYSQAS 59
Db 966 EDKISLLKEQMFNLNLDLQKGMKEK--EKADFKKISILQNNKEVEAVKSEYKSL 1023
QY 60 VLVDGIKVLMSDQKYFEATQTVYEWGCVVTVLLSAYILLDFEYNEKASAKDI---- 115
Db 1024 KIQND-----LDQOTIYANTQNNYE-----QELQKHADVSKTISELR 1061
QY 116 -LIRILDGVKKLN-----EAQKSLTTSQSFNNAAGKLLA---LDSOLTNDPSEKSYFQ 167
Db 1062 EQHTYKGVQKVTNLNLRDQLENALKENKESWSQKESLLEQLDLSNRIEDLSQNKLLY 1121
QY 168 SQVDRIRKEAYAGAAGIVAGPFGLLISYSTAAGVIEGKLIPELNNRLKTVQNFFTLSA 227
Db 1122 DQI-----QIYTAADKEVNNSTNG-----PGLNNILITLRERDILT 1159
QY 228 TVQKANKDIDAAGKLAETAAIGETKTETTR-----FYVDYDDL-----LSLL 274
Db 1160 KTVVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNLL 1219
QY 275 KGAAKKMINTCNEYQQRHGKK 295
Db 1220 R---ESNITLRNELENNNKK 1237

RESULT 36
ABP39975
ID ABP39975 standard; protein; 1211 AA.

XX ABP39975;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4820.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection; antibacterial; gene therapy.

XX

```
OS Staphylococcus epidermidis.
XX US6380370-B1.
XX 30-APR-2002.
XX 13-AUG-1998; 98US-00134001.
XX 14-AUG-1997; 97US-0055779P.
XX 08-NOV-1997; 97US-0064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDE; ABN92520.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
XX polypeptide, useful for diagnosing and treating bacterial infections.
XX Disclosure; SEQ ID NO 4820; 267pp; English.
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences can
XX also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life cycle
XX or inhibit S. epidermidis infection. N.B. The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from the USPTO web site
XX
SQ Sequence 1211 AA;
Query Match 7.1%; Score 107; DB 5; Length 1211;
Best Local Similarity 18.3%; Pred. No. 13;
Matches 61; Conservative 59; Mismatches 131; Indels 82; Gaps 9;
QY 7 EQTVVVK-----SAIETADGALDLYNKYLDQVDPWPKTFDETIKLSRFKQYSQASV 60
DB 185 EESAGVLKYKKRAESIQKLDHTEDNLNRVEDILYDLEGRVPLKEAAIAKEYKQLSKE 244
QY 61 LVGDIKVLMDSQKYPEATQTVVWC-----GVVTQLLSAYILLFDBYNE 106
DB 245 MEQSDVITVTSIDHYTEDNQRDLERLNHLKSQAQKEGQAQINQLLQKY----- 295
QY 107 KKASAKDILIRILDGCKVLEAKSLLTSSQSFNNASGKLLALDSQLTND-----FS 160
DB 296 -KGRQON-----DYDIEKLN---YELVKATENYEQLSGKLVLEERKKNQSETNARYE 345
QY 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLIITSYIAAGVIEGKLIPELNRL-----K 216
DB 346 BELDNLSEQSDISKNEKAQNE-----KLLADLKNKQKQLNK 381
QY 217 TVQNFFTSLSATVKQANKIDAAKVLKATIAAIGETETTRFFVYVDVDMLSLKG 276
DB 382 EVQELESLLYISDQHKLEIEIKNSYYITLMSQSVVNNDRIFLEHTINENEAKSRIDS 441
QY 277 AAKGMINTCNEYQO-----RHGKKTLPFV 300
DB 442 RLVEAFNQLKDIOQNTITQTOKEYSKSKMEKV 474
RESULT 37
ABB49720
ID ABB49720 standard; protein; 927 AA.
XX AC ABB49720;
XX DT 05-FEB-2002 (first entry)
XX
```

```
DE Listeria monocytogenes protein #2424.
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX Listeria monocytogenes.
XX WO200177335-A2.
XX 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR001118.
XX 11-APR-2000; 2000FR-00004629.
XX (INSP ) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX Daniels J, Gobbel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;
XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and related
XX polypeptides.
XX Claim 6; SEQ ID NO 2425; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccine compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 927 AA;
Query Match 7.0%; Score 106.5; DB 5; Length 927;
Best Local Similarity 20.5%; Pred. No. 9.8;
Matches 78; Conservative 44; Mismatches 131; Indels 127; Gaps 16;
QY 1 MTSIFAEQTVVVKSAIETADG---ALDLYNKYLDQVDPW---KTFDETIKLSRFKQE 53
DB 160 VTKSYAEATFDKIK---ESGDGFAQADGSGKIKDGLVKSQEGNKTIKTLKTLADSSLT 216
QY 54 YSQEASVL-----VGDIK-----VLLMDS 72
DB 217 FKDGANTLEVGLTYTDGVNNTAAAGDKLNAGVSTLAAGVPLKDGVAALDGGATKLASG 276
QY 73 QDKYFEATQTVVWCQGVVTQLLSAYILLFDBYNEKKASAKDILIRILDGCKVLEAKQ 132
DB 277 VSTYTSQVDTL---AGGINQAYTGSTALSDGLNKNMGS-----VPTLASGITQLNNGQK 327
QY 133 SLITSSQSFNNASGKLLA---LDSQLTN-----DFSEKSSYFQSDVRIRKEAYAGAA 182
```

328 SLATGLDLSVDSGNKLSAGLKELDGNLTKQKIAQLKQGMNDLQGLDQLNQSVNGEDA 387
 183 AGIVAGPFGLLIISYSTAAGVIEGKLPENLRL- - - - -KTQVQFFTSLSAT 228
 388 A- - - - -LAKQLAT- - - - -LQKSLDQLQGLTFIKSNANFDAEAIKSKINATAGVSAE 434
 229 VKQ- - - - -ANKDIDAAKLKLATIAAIGET- - - - -ETETRYVDYDMLSL 273
 435 DKQIIDAQADLDKETQKSAQTQVATVEQLSGLSLDLAAITQV- - - - -EL 483
 274 LKGAAKKQVINTCNEYQORHG 293
 484 QTGVAKISAG- - - - -YQAVHG 499

RESULT 38
 AAW22775
 ID AAW22775 standard; protein; 1312 AA.
 XX AC AAW22775;
 XX DT 21-DEC-1998 (first entry)
 XX DE Human RAD50.
 XX KW Human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
 KW central nervous system.
 XX OS Homo sapiens.
 XX PN WO9727284-A2.
 XX PD 31-JUL-1997.
 XX PF 24-JAN-1997; 97WO-US001299.
 XX PR 26-JAN-1996; 96US-00592126.
 PR 17-JUL-1996; 96US-00687080.
 XX PA (GENE-) GENE LABS TECHNOLOGIES INC.
 XX PI Dolganov G;
 XX PS WPI; 1997-393672/36.
 DR N-PSDB; AAT75237.
 XX PT Human tumour suppressor gene RAD50 - useful to detect predisposition to,
 PT decrease risk of and treat cancer, also Septin-2 homologues.
 XX PS Claim 5; Page 82-86; 195pp; English.
 XX CC The human RAD50 (hRAD50) is involved in DNA repair and has tumour
 CC suppression activity, can be used to detect predisposition to, decrease
 CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
 CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
 CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
 CC anaemia with excess blasts. Also disclosed in this invention is Human
 CC Septin-2 homologues of which may be used as targets for cancer therapies
 CC and central nervous system directed treatment methods, and to measure the
 CC proliferative potential of selected cell types
 XX SQ Sequence 1312 AA;
 Query Match 7.0%; Score 106.5; DB 2; Length 1312;
 Best Local Similarity 18.9%; Pred. No. 15;
 Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
 OY 10 VEVKSAJETADGALDLYNKYLDQVTPKTFDEIK- - - - -ELSRFKQY 54
 DB 502 MEVISLQNEKAD- - - - -LDRLRLKLDQEMEQLNHHHTTQMEMLTKQKADKDEQIRKISRH 559
 OY 55 SOEASVLGDIKVLMLDSQDKYFEATQTVYEWCGVTTQLLSAYILLFDEYNEKASAKQD 114

560 SDELTSLG- - - - -YFPNKKQLEDWLHKS- - - - -EINQTR- - - - - 591
 115 ILTRILDDGVKLINEAQKILLTSSQSFNNASGKLALDSQLTNDPSEK- - - - -SSYFQS 168
 592 - - - - -DRUAKLN- - - - -KELASSEQNKHNINNELKKEQLSS-YEDKLFVCGSQDFES 640
 169 QVDRIKRE- - - - -AVAGAAAGIVAGPF- - - - - 190
 641 DLDRLEKEIEKSKQRAMLAGATA- - - - -VYSQFITQLTENOSCCPVCQRFQTEAEQLQEV 698
 191 - - - - -GLI- - - - -ISYSIAAGVIEGKLPENLRLKT 217
 699 SDLOSRLAPDKLSTESLKKKKERDEMLGLVPMRQSIID- - - - -LKEKEIPELRNKLQ 756
 218 VONEFTSLSATVQANKDIDAAKLKLATEIAAIGETETET- - - - -RFVVDY 266
 757 V- - - - -NRDIQRLKNDIBEQETLLGTIMPEESAKVCLTDVTIMERFQML 802
 267 DDLMLSLKGA- - - - -KMINTCNEYQORHGK 295
 803 KDVETKIAQAALQGLDLDRTVQOVNQEKQKQHK 838

RESULT 39
 AAW71295
 ID AAW71295 standard; protein; 1312 AA.
 XX AC AAW71295;
 XX DT 25-NOV-1998 (first entry)
 XX DE Human homologue of yeast RAD50.
 XX KW Human homologue; Yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;
 KW immunomodulatory activity; identification; activated T-cell.
 XX OS Homo sapiens.
 XX PN WO9838306-A1.
 XX PD 03-SEP-1998.
 XX PF 27-FEB-1997; 97WO-US003159.
 XX PR 27-FEB-1997; 97WO-US003159.
 XX PA (GENE-) GENE LABS TECHNOLOGIES INC.
 XX PI Dolganov G;
 XX PS WPI; 1998-481207/41.
 DR N-PSDB; AAV59979.
 XX CC Novel human immunomodulatory poly-peptide(s) - have homology to the yeast
 CC RAD50 or Drosophila Septin-2 proteins.
 XX PS Disclosure; Page 136-140; 155pp; English.
 XX CC The present sequence represents a human homologue of the yeast S.
 CC cerevisiae gene RAD50. The present sequence has 35% overall homology to
 CC the yeast RAD50 gene, and is expressed in activated T-cells, testis,
 CC foetal liver and heart tissues. The specification also describes
 CC sequences encoding human homologues of the yeast RAD50, the Drosophila
 CC Septin-2 and Acyl-CoA synthetase. The proteins have immunomodulatory
 CC activity. The nucleic acids and proteins can be used to identify
 CC activated T-cells in a sample population. They can also be used to
 CC isolate and identify sequences encoding other proteins or other compounds
 CC having immunomodulatory activity
 XX SQ Sequence 1312 AA;
 Query Match 7.0%; Score 106.5; DB 2; Length 1312;
 Best Local Similarity 18.9%; Pred. No. 15;

CC sequences which encode them. In particular, the genes are derived from
CC Lactobacillus johnsonii. The invention may allow a better understanding
CC of the mechanism by which probiotics (such as L. johnsonii) can bind to
CC mucins, large glycoproteins present on the surface of a large number of
CC epithelial cells including those in the gastrointestinal tract, the lung
CC or the uterine cervix. The present sequence is that of an L. johnsonii
CC mucin binding protein of the invention.
XX
XX
SQ Sequence 4734 AA;

Query Match 7.0%; Score 106.5; DB 8; Length 4734;
Best Local Similarity 23.8%; Pred. No. 81;
Matches 73; Conservative 38; Mismatches 109; Indels 87; Gaps 13;
QY 2 TSIAEQVTEVVKSAIETAGALDLYN-----KYLQV-----IPWK 38
Db 4126 TNISSEKTDLIKQATEAANKNNATNTSEVETAQVDGEKAIADVTGPGLSDIKKE 4185
QY 39 TFDETIKELSRFKQEYSQASVLDGDKVLLMSQDKYFEATQTVYEWCGVVTQLLSAYI 98
Db 4186 SIDLINKALNE-KQDEINNASNLQDESTELIDQAKKI--ATEAINEINNAQTN----- 4236
QY 99 LLEFDEYNKSAQKDI---LIRILDGQVKKLENAQKSLTSSOS-FNNASGKLLALDSQ 154
Db 4237 ---DEAKAADTVGVKHNVSIPSTEDAKKNATQIIDDALNSKQNEINNASN---LTDSE 4290
QY 155 LTNDSEKSYFQSQVDRIKKEAVAGAAAGIVAGPFGLLIISYSTAAGVIEGKLIPELNNR 214
Db 4291 KTDLINQ-----ATEIANAADKAINSATT-----NTAVEAAEYKGVADINN- 4331
QY 215 LKTQNFSTSLGATVQKQKDI-----DAAKKLK-----ATEIAAIGEII 253
Db 4332 -----IHFTNLDDSKKAANSALIEDALTTKKDEINNASNLSDSEKAKLINQATEIANAKA 4386
QY 254 KTEETT 260
Db 4387 AINNATT 4393

RESULT 42
ABB71136
ID ABB71136 standard; protein; 7201 AA.
XX ABB71136;
XX
XX
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 40200.
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL15239.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
PT

XX Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 7201 AA;

Query Match 7.0%; Score 106.5; DB 4; Length 7201;
Best Local Similarity 18.9%; Pred. No. 1.4e+02;
Matches 68; Conservative 51; Mismatches 131; Indels 109; Gaps 14;
QY 18 ETADGA-LDLYNKYLDQVIPW-----KTDETIKEL 47
Db 3504 QEAQGVQLSSYQDILNQTVMWLDQVEKLIHNPASWTSQAETRSKLYKYKATNQDINSH 3563
QY 48 SRFKQEYSQASVLDGDKVLLMSQDKYFEATQTVYEWCG-----VVTQLLSAYILL-- 100
Db 3564 KRIVEAVNEKAAALIGSAAAPANADEISKAVAEVKNRYDQVQDCAKLVADLDGAFVYQQ 3623
QY 101 PDEYNKSAQKQIILRI-----LDGQVKKLENAQKSLTSSOSFNNASGKL 148
Db 3624 FSELOKAQDYQKMLWRLTGTVDYSGNKAALQARLQKINEIODAL-----PEGVAKL 3676
QY 149 LALDSQLTNDPSEKSYFQSQVDRIKKEAVAGAAAGIVAG--PFGLLIISYSTAAGVIEGK 206
Db 3677 KSLEDHIEQQ-----ASNIPARSKVEMARDLANLHADFEKFGASLS-DVKSG-LENR 3726
QY 207 L-----IPELNNRLKTVQ-----NFTSLSATVQKQKQIDA 238
Db 3727 LQWNDYEINLDRLLITWLGEAENSLKYNLKSFFKEEQLNGFQSLAQLNRQNEADFDK 3786
QY 239 AKLKATEIAAIGB-----IKTETETFRFVYDYLMLSLKGAACKMINTCNEYQQRHG 293
Db 3787 VKDDTSELVQSGGTRIYVNVQVSSRF-----QSIQATAKEILLKKEQAVQDHG 3836
RESULT 43
ABB64018
ID ABB64018 standard; protein; 685 AA.
XX ABB64018;
XX
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 18846.
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX

12-JUN-2003.
 25-NOV-2002; 2002WO-US036123.
 29-NOV-2001; 2001US-0333777P.
 18-NOV-2002; 2002US-0426742P.
 (AMHP) WYETH HOLDINGS CORP.
 Fletcher LD, Memichael JC, Russell DP, Zagursky RJ;
 WPI: 2003-505284/47.
 N-PSDB; ADB08929.
 New Alloiooccus otitidis polynucleotides and polypeptides, useful for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials.
 Claim 33; SEQ ID NO 2870; 1019pp; English.
 The present invention describes an isolated polynucleotide (I) of Alloiooccus otitidis genomic DNA, which encodes an antigenic protein. Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1) an isolated polypeptide that is encoded by the polynucleotide (I); (2) an expression vector comprising the novel isolated polynucleotide (I), its complement, degenerate variant or fragment; (3) a genetically engineered host cell, transfected, transformed or infected with the vector of (2); (4) an antibody specific for the polypeptide of (1); (5) an immunogenic composition comprising the polypeptide, its complement, biological composition comprising the polynucleotide that is comprised in the equivalent or fragment, or the polynucleotide that is comprised in the expression vector; (6) a pharmaceutical composition comprising the polypeptide of (1) and a carrier; (7) a protein chip comprising an array of the polypeptides of (1), their biological equivalent or fragment; (8) immunising against Alloiooccus otitidis by administering to a host the immunogenic composition; (9) detecting and/or identifying Alloiooccus otitidis in the biological sample; (10) a kit comprising a container containing the novel polynucleotide, its degenerate variant or fragment, or the antibody of (4); and (11) producing a polypeptide by culturing the genetically engineered host cell under conditions suitable to produce the polypeptide from the culture. (I) can be used in gene therapy. The polynucleotides, polypeptides, antibodies and compositions of the present invention can be used for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials. The polynucleotides are useful for expressing and detecting Alloiooccus otitidis. The present sequence represents an Alloiooccus otitidis antigen protein from the present invention.

Query Match 7.0%; Score 106; DB 6; Length 863;
 Best Local Similarity 20.3%; Pred. No. 9.9;
 Matches 61; Conservative 54; Mismatches 107; Indels 78; Gaps 11;
 4 IFABQTVVVKSAIETA-----DGLDLYNKYLD-----Q 33
 460 MFSREDMEIQOLIKSINFELFENQMVNNHNSNEKDYLYLYKELIEATSELFWD 519
 34 VIPWKTDETIKLSR-----FKQYEQASVVLGDIK--VLLMDSQKYP 77
 520 TITYKSNKTEITELIKVPSITYAGSGICLFLEYAVENLLDKLHLMKDIYRSFI 579
 78 BATQTVYEWCVVQLSAYITLLPDEYNEKASAKQDILRLDDGVKLNKAEKSLTS 137
 580 KIVNNPYSCGIYDGTSGFLYLIY-KYQVYTRFNKNKELLISFLKIMTKVQYDKQDIMS 638
 138 SQSFNAGKLLALDSQLTNDFSEKSSYFQSOV-----DRIRK-----EAYAGAAG 184
 639 GQ-----AGLKLILLYLN--GEKYDWIKEGVINQIIIDKINSNVLHKKYLYGSHG 691
 185 IVAGPFGLLIISYIAAGVIEGKLPILNLRUKTVQNTFTSLSATVQKANKDIDAAKLKA 244
 692 LV-----GIISVLVESTKINKIFKNIENIVENFVYLYLS-SVMENKGIPTIEKLS 740

RESULT 46
 AAG82283
 ID AAG82283 standard; protein; 885 AA.
 XX
 AC AAG82283;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1660.
 XX
 KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
 XX
 KW endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US030782.
 XX
 PR 09-NOV-1999; 99US-0164258P.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 XX Kimmerly WJ;
 XX
 DR WPI: 2001-316495/33.
 XX
 DR N-PSDB; AAH53133.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis.
 XX
 PS Claim 18; Page 462; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II) given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I
 CC and (II) can have antibacterial activity and therefore can be used in
 CC vaccination. The nucleic acids (I) may be used to produce the S.
 CC epidermidis polypeptides (II) via the production of vectors containing
 CC them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to AAH5098
 CC represent oligonucleotide sequences and primers which are used in the
 CC exemplification of the present invention. N.B. The present invention
 CC specifically claims all the polynucleotide sequences given in the
 CC sequence listing of the present specification, however the sequence
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
 CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 CC for SEQ ID NO:4455 to 4464
 XX
 SQ Sequence 885 AA;
 Query Match 7.0%; Score 106; DB 4; Length 885;
 Best Local Similarity 18.5%; Pred. No. 10;
 Matches 54; Conservative 51; Mismatches 111; Indels 76; Gaps 8;
 42 ETIKELSRFKQYEQASVVLGDIKVLMDSQKYPFATQTVYEW 87
 2 EPLKEAAIAKQYKQSKEMEQSDVITVSDIDHYEDNQLRLNHLKSAQAEKGGQ 61
 88 GVTQTLISAYILLPDEYNEKASAKQDILRLDDGVKLNKAEKSLTSQSQFNNAAGK 147
 62 AQINLLQKY-----KGKQON-----DYDIEKLN---YELVKATENYEQLSGK 102
 148 LLALDSQLTND-----FSEKSSYFQSOVDRIKAEYAGAAAGVAGPFGLLIISYIAAG 201

CC The specification describes a method for directing integration of a
CC nucleic acid of interest to a pre-determined site, where the nucleic acid
CC has homology at or around the pre-determined site, in a eukaryote with a
CC preference for non-homologous recombination. The method comprises
CC steering an integration pathway towards homologous recombination. The
CC method is useful for directing integration of a nucleic acid of interest
CC to a subtelomeric and/or telomeric region in an eukaryote with a
CC preference for non-homologous recombination. The nucleic acid of interest
CC comprises an inactive gene to replace an active gene, or vice versa, is a
CC portion of a gene delivery vehicle, confers a desired property to the
CC eukaryotic cell, or encodes a therapeutic proteinaceous substance or a
CC substance conferring resistance for an antibiotic substance to a cell.
CC The method is useful for improving gene targeting efficiency. The method
CC is useful in the replacement of an active gene by an inactive gene, for
CC e.g. for the inactivation of genes controlling undesired side branches of
CC metabolic pathways, to increase the quality of bulk products such as
CC starch, or to increase the production of specific secondary metabolites
CC or to inhibit formation of unwanted metabolites, and also to inactivate
CC genes controlling senescence in fruits and flowers or that determine
CC flower pigments. The method is also useful for replacing an inactive gene
CC by an active gene. For e.g. the replacement of a defective p53 by an
CC intact p53. Many tumours acquire a mutation in p53 during their
CC development which renders it inactive and often correlates with a poor
CC response to cancer therapy. By replacing the defect p53 by an intact p53,
CC e.g. through gene therapy, conventional anti cancer therapy have better
CC changes of succeeding. The method is also useful for therapeutic
CC proteinaceous substance integration. A tumoricidal gene can be delivered
CC to a pre-determined site present only in e.g. proliferating cells, or
CC present only in tumour cells, e.g. to the site where a tumour antigen is
CC expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved
CC in non-homologous recombination

XX Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 5; Length 1318;
Best Local Similarity 18.4%; Pred. No. 19;
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;
QY 10 VEVVKSATETADGALDLYNKYLDQVTPWKTFDETIK-----ELSRFKQY 54
DB 508 MEVISLQNEKAD--LDRTLRLKLDQEMEQLNHHHTTTRTQMEMLTKADKADQEKIRKSRH 565
QY 55 SQEASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFPDYEYNEKKASAKQD 114
DB 566 SDELTSLLG-----YFPNKKQLEDWLHSHSK-----EINQTR----- 597
QY 115 ILIRILDGVKKLNEAOKSLTSSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168
DB 598 -----DRLAKLN---KELASSEQNKHNINNELERKEQQLSS--YEDKLFVDCGSQDFES 646
QY 169 QVDRIRKE-----AVAGAAA-----GIVAGPFLGIISYSIAAGV 202
DB 647 DLDRLKEEIEKSKQRAMLAGATVYSQFITQLTDENQSCCPVCQVQTEAELQEAISD 706
QY 203 IEGKL-----IPELNNRLKTVQNF 221
DB 707 LQSKRLAPDKLSTESLKKXKRRDEMGLAPMRQSIIDLKEKEIPELRNKLQNV--- 763
QY 222 FTLSLTVQANKDIDAAKLKATEIAAIGETKTETTT-----RFYVDYDDIM 270
DB 764 -----NRDIOQLRNDIEEQETLLGTTIMPEESAKVCLTVDVTIMERQOMELKDVE 812
QY 271 LSLKGAAK-----KMINTCNEYQQRHGKK 295
DB 813 RKTAQQAQKLGIDLDRLTRVQVNOEKQEKQHK 844

RESULT 49
ADJ68860
ID ADJ68860 standard; protein; 1318 AA.

XX
AC ADJ68860;

DT 06-MAY-2004 (first entry)
XX Human heat mitochondrial protein as a therapeutic target SeqID666.
DE
XX
XX Mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis; LHON;
KW Leber's hereditary optic neuropathy; LHON;
KW Mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.
OS Homo sapiens.
XX
XX WO2003087768-A2.
PN
XX
XX 23-OCT-2003.
PD
XX
XX 04-APR-2003; 2003WO-US010870.
PF
XX
XX 12-APR-2002; 2002US-0372843P.
PR
XX
XX 17-JUN-2002; 2002US-0389987P.
PR
XX
XX 20-SEP-2002; 2002US-0412418P.
PR
XX
XX (MITO-) MITOKOR.
PA
XX
XX (BUCK-) BUCK INST AGE RES.
PA
XX
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
PI
XX
XX WPI; 2003-845369/78.
DR
XX
XX Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
PT
XX
XX Claim 1; SEQ ID NO 666; 180pp; English.
PS
XX
XX This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nootropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
XX
XX Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 7; Length 1318;
Best Local Similarity 18.4%; Pred. No. 19;
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;
QY 10 VEVVKSATETADGALDLYNKYLDQVTPWKTFDETIK-----ELSRFKQY 54
DB 508 MEVISLQNEKAD--LDRTLRLKLDQEMEQLNHHHTTTRTQMEMLTKADKADQEKIRKSRH 565
QY 55 SQEASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFPDYEYNEKKASAKQD 114
DB 566 SDELTSLLG-----YFPNKKQLEDWLHSHSK-----EINQTR----- 597
QY 115 ILIRILDGVKKLNEAOKSLTSSQSFNNASGKLALDLSQLTNDFSEK-----SSYFQS 168
DB 598 -----DRLAKLN---KELASSEQNKHNINNELERKEQQLSS--YEDKLFVDCGSQDFES 646

QY 169 QVDRIRKE-----AYAGAA-----GIVAGPFGLLISYSIAAGV 202
Db 647 DLDRKEIEKSKQRAMLAGATAVYSQFTQLTDENQSCCPVCQVQFQTEAELOEAI 706
QY 203 IEGKL-----IPELNNRLKTYQNF 221
Db 707 LQSKRLAPDKLKSTESSELKKKRRDEMGLAPMQSIIDLKEKEIPELNRKLNQV--- 763
QY 222 FTLSATVQKANKDIDAAKLKLATEIAAIGETETETTT-----RFVYDDLM 270
Db 764 -----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQMKDVE 812
QY 271 LSLKGAAG-----KMINTCNEYQORHGKK 295
Db 813 RKTAQQAQAKLQGLDLDRTVQOVNQEKQKH 844

RESULT 50

ADJ66509
ID ADJ66509 standard; protein; 1318 AA.
XX AC ADJ66509;
XX DT 06-MAY-2004 (first entry)
XX DE RAD50 homolog HSRAD50 for anti-cancer protein complex.
XX KW neuroprotective; cytostatic; gene therapy; protein complex;
XX KW cellular network; cancer; neurodegenerative disease; drug target.
XX OS Homo sapiens.
XX PN WO2004009622-A2.
XX PD 29-JAN-2004.
XX PF 18-JUL-2003; 2003WO-EP007835.
XX PR 19-JUL-2002; 2002EP-00016109.
XX PR 19-JUL-2002; 2002EP-00016111.
XX PR 19-JUL-2002; 2002EP-00016123.
XX PR 19-JUL-2002; 2002EP-00016128.
XX PR 22-JUL-2002; 2002EP-00016427.
XX PA (CELL-) CELLZONE AG.
XX PI Marino A, Bouwmeester T, Bauer A, Drewes G, Marzioch M, Kruse U;
XX PI Superti-Furga G, Eberhard D, Ruffner H, Hobson S, Helftenbein G;
XX PI Cruciat C;
XX DR WPI; 2004-123372/12.

XX New protein complexes of cellular networks underlying the development of
XX cancer and other diseases, useful for diagnosing and/or treating
XX neurodegenerative diseases or cancer, and in drug screening.

XX Disclosure; SEQ ID NO 39; 809pp; English.

XX The invention relates to a protein complex of cellular networks
XX underlying the development of cancer and other diseases. The complex (I)
XX comprises at least one first and second proteins selected from any of the
XX proteins listed in the specification, or their functionally active
XX derivatives, fragments, homologues or variants, the variants being
XX encoded by a nucleic acid that hybridizes to the nucleic acid encoding
XX the protein under low stringency conditions. A complex (II) comprises at
XX least two of the second proteins, where the low stringency conditions
XX comprise hybridization in a buffer comprising 35% formamide, 5 x SSC, 50
XX mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% BSA, 100 microg/ml
XX denatured salmon sperm DNA, and 10% (wt/vol) dextran sulfate for 18-20
XX hours at 40 deg C, washing in a buffer consisting of 2 x SSC, 25 mM Tris-
XX HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS for 1.5 hours at 55 deg C, and
XX washing in a buffer consisting of 2 x SSC, 25 mM Tris-HCl (pH 7.4), 5 mM
XX EDTA, and 0.1% SDS for 1.5 hours at 60 deg C. The composition and methods

CC are useful in diagnosing or treating diseases and disorders, preferably
CC neurodegenerative diseases. These may also be used as a drug target or in
CC manufacturing a medicament for the treatment or prevention of the above-
CC mentioned diseases or disorders. The composition may also be used for
CC treating cancer. This sequence represents one of the proteins of the
CC complex of the invention.

XX Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 8; Length 1318;
Best Local Similarity 18.4%; Pred. No. 19;
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;

QY 10 VEYVKSATETADGALDLYNKYLDOVIPKTFDTETIK-----ELSRFQOEY 54
Db 508 MEVISLQNEKAD--LDRTLKLDQEMQLNHHHTTTTQEMMLTKDKADKDEQIKKISRH 565
QY 55 SQEASVLVGBDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAVILLFDEYNEKKASAKQD 114
Db 566 SDELTSLLG-----YFPNKKQLEDWLHLSKSK-----EINQTR----- 597
QY 115 ILIRIILDDGVKLNBAQKSLTSSQSPNNSAGKLALDLSQLTNDNFSEK-----SSYFQS 168
Db 598 -----DRLAKLN--KELASSEQNKHNINNELERKEEQLSS-YEDKLPDVCQSQDFES 646
QY 169 QVDRIRKE-----AYAGAAA-----GIVAGPFGLLISYSIAAGV 202
Db 647 DLDRKEIEKSKQRAMLAGATAVYSQFTQLTDENQSCCPVCQVQFQTEAELOEAI 706
QY 203 IEGKL-----IPELNNRLKTYQNF 221
Db 707 LQSKRLAPDKLKSTESSELKKKRRDEMGLAPMQSIIDLKEKEIPELNRKLNQV--- 763
QY 222 FTLSATVQKANKDIDAAKLKLATEIAAIGETETETTT-----RFVYDDLM 270
Db 764 -----NRDIQRLKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQMKDVE 812
QY 271 LSLKGAAG-----KMINTCNEYQORHGKK 295
Db 813 RKTAQQAQAKLQGLDLDRTVQOVNQEKQKH 844

Search completed: January 5, 2005, 10:56:28

Job time : 70.7089 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:51:49 ; Search time 16.0526 Seconds
(without alignments)
1260.042 Million cell updates/sec

Title: US-09-993-292B-2
Perfect score: 1515
Sequence: 1 MTSIPAEQTVVVKSALETA.....NEYQQRHGKTLFEPDVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1396	92.1	309	1	Sequence 3, Appli
2	1396	92.1	309	5	Sequence 3, Appli
3	121.5	8.0	1036	4	Sequence 7736, Ap
4	114	7.5	515	4	Sequence 5317, Ap
5	112.5	7.4	722	4	Sequence 20613, A
6	111	7.3	1231	4	Sequence 5150, Ap
7	108.5	7.2	1066	3	Sequence 8, Appli
8	108.5	7.2	1066	4	Sequence 8, Appli
9	108.5	7.2	1066	4	Sequence 8, Appli
10	108.5	7.2	3696	3	Sequence 5080, Ap
11	107	7.1	1211	3	Sequence 4820, Ap
12	106.5	7.0	1312	2	Sequence 148, App
13	106.5	7.0	1312	2	Sequence 51, Appl
14	106.5	7.0	1312	4	Sequence 148, App
15	106	7.0	815	4	Sequence 4284, Ap
16	106	7.0	885	4	Sequence 1660, Ap
17	105.5	7.0	924	4	Sequence 18798, A
18	104.5	6.9	566	4	Sequence 807, App
19	104.5	6.9	1196	4	Sequence 3944, Ap
20	103.5	6.8	718	4	Sequence 2753, Ap
21	103	6.8	606	4	Sequence 2, Appli
22	103	6.8	631	4	Sequence 11, Appl
23	102	6.7	1976	4	Sequence 1078, Ap
24	101	6.7	477	1	Sequence 3, Appli
25	101	6.7	477	1	Sequence 3, Appli
26	101	6.7	477	3	Sequence 3, Appli
27	101	6.7	1939	3	Sequence 1, Appli

Query Match 92.1% Score 1396; DB 1; Length 309;

ALIGNMENTS

RESULT 1

US-08-557-115-3
; Sequence 3, Application US/08557115
; Patent No. 5731151
; GENERAL INFORMATION:
; APPLICANT: King, Harold C.
; APPLICANT: Sathish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greene, Jamie L.
REGISTRATION NUMBER: 32,467
REFERENCE/DOCKET NUMBER: 03063-0171US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 818-3700
TELEFAX: (404) 818-3799
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-557-115-3

Sequence 917, App
Sequence 844, App
Sequence 5793, App
Sequence 1077, App
Sequence 936, App
Sequence 3159, App
Sequence 5178, App
Sequence 3594, App
Sequence 5599, App
Sequence 3, Appli
Sequence 170, App
Sequence 170, App
Sequence 16013, A
Sequence 91, Appl
Sequence 1084, App
Sequence 2, Appli
Sequence 6, Appli
Sequence 1154, App
Sequence 1, Appli
Sequence 1, Appli
Sequence 66, Appl
Sequence 8352, App
Sequence 6, Appli

Best Local Similarity 91.4%; Pred. No. 3e-130; Matches 275; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MTSIFABQTVVVKSAIETADGALDLYNKYLDQVDPKTFDETIKELSRFKQESQASV 60
Db 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVDPKTFDETIKELSRFKQESQASV 60

Qy 61 LVGDIKVLMDSDQKYPEATQTVYEWCGVATQLLAAYILLFDBYNEKKSAAQKDILIRIL 120
Db 61 LVGDIKVLMDSDQKYPEATQTVYEWCGVATQLLAAYILLFDBYNEKKSAAQKDILIRIL 120

Qy 121 DDGKKLNEAQSLLTSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180
Db 121 DDGTTKLEAQSLLVSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180

Qy 181 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNRLKTVQNFFTSLATVQKANKDIDAAK 240
Db 181 AAAGVAGPFGLLIISYIAAGVIEGKLIPELNRLKTVQNFFTSLATVQKANKDIDAAK 240

Qy 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQKHGKTLFEV 300
Db 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQKHGKTLFEV 300

Qy 301 P 301
Db 301 P 301

RESULT 2
PCT-US94-05869-3
; Sequence 3, Application PC/TUS9405869
; GENERAL INFORMATION:
; APPLICANT: King, C. H.
; APPLICANT: Satish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIS
; TITLE OF INVENTION: OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 1200, The Candler Building
; STREET: 127 Peachtree Street, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05869
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US94-05869-3

Query Match 92.1%; Score 1396; DB 5; Length 309;
Best Local Similarity 91.4%; Pred. No. 3e-130;

Matches 275; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MTSIFABQTVVVKSAIETADGALDLYNKYLDQVDPKTFDETIKELSRFKQESQASV 60
Db 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVDPKTFDETIKELSRFKQESQASV 60

Qy 61 LVGDIKVLMDSDQKYPEATQTVYEWCGVATQLLAAYILLFDBYNEKKSAAQKDILIRIL 120
Db 61 LVGDIKVLMDSDQKYPEATQTVYEWCGVATQLLAAYILLFDBYNEKKSAAQKDILIRIL 120

Qy 121 DDGKKLNEAQSLLTSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180
Db 121 DDGTTKLEAQSLLVSSQSFNNASGKLLALDQSLTNDPSEKSSYFQSOVDRIKREAYAG 180

Qy 181 AAAGIVAGPFGLLIISYIAAGVIEGKLIPELNRLKTVQNFFTSLATVQKANKDIDAAK 240
Db 181 AAAGVAGPFGLLIISYIAAGVIEGKLIPELNRLKTVQNFFTSLATVQKANKDIDAAK 240

Qy 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQKHGKTLFEV 300
Db 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQKHGKTLFEV 300

Qy 301 P 301
Db 301 P 301

RESULT 3
US-09-543-681A-7736
; Sequence 7736, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7736
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7736

Query Match 8.0%; Score 121.5; DB 4; Length 1036;
Best Local Similarity 20.7%; Pred. No. 0.0068;
Matches 80; Conservative 66; Mismatches 126; Indels 115; Gaps 17;

Qy 2 TSIFAEQTVBWK-----SAIETADGALDLYNKYLDQVDPKTFDETIKELSRFKQ 53
Db 167 TALSSMEINELIKKQNGENVSSSELAKASIDLINQLVD-----TASSLNNNISAFSQ 220

Qy 54 YSQEASV-----VGDIKVLLMSQDKYFEATQTVYEWCGVATQLLAAYILL-FDEY 104
Db 221 LNKLSVLSNTKHLNGVGN-KLQNLPLNDKLTGLDVT-----SGLSAISASFILSNADAD 276

Qy 105 NEKASAAQKDLIRILDGVKKLNE-----AOKSLTSSQSFNNASGKLL 149
Db 277 TGTAAAGVELTITKLVGNVGVKANSQYILAQORVAGLSTSAASAGLIASAVTLAISPLSFL 336

Qy 150 ALDSQL--TNDPSEKSSYFQ-----SQVDRIKREA-----YAGAAA 183
Db 337 AIADQFKANKIEEVSQRFKFGYEGSLLAAPFKETGAIDASTTINTALGTISAGISA 396

Qy 184 GIVAGPFGLLIISYIAA--GVIEG-----KLPELNRLKTV-----QNFTS- 224
Db 397 ASTASLIGAPISALVGAITGIIISGILEASQSMFHVANRWANTIAEWKTHGNFPENG 456

Qy 225 -----LSATVK---QANKDIDAAKIKLAT-----EIAAIGETETETTRFYV 264

[illegible]

Qy	176	----	EAVAGAAAGIVAGPFGLLISYSIAAGVIEGKLIPELNRLRKTQNVFTTSATVKQA	232
Dd	230	RGYEYIRSQGEVRDG-----	VTVNAAAVVPEVIGEVFDLKR--NYNLAQYATVKTV	282
Qy	233	NKDIDAAKLKLAT-----	EIAAIGEIKTETETTFYVDY-----	266
Dd	283	SN--GQCKYPVATNQOAVLATKAEIAEIGDIDAEFMS---	VDYKVETRAGKIALSNEVV	337
Qy	267	DLMLLSLLKGA-----	AKGMINTCNEY-----QQRHGKKTLFEVPDV	303
Dd	338	EDSAVNIVEGVQDLAKLVENTONKHIMDLLTKTTFKTAATLDL	382	
 RESULT 5 US-09-248-796A-20613 ; Sequence 20613, Application US/09248796A ; Patent No. 6747137 ; GENERAL INFORMATION: ; APPLICANT: Keith Weinstock et al ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDID ; FILE REFERENCE: 107196.132 ; CURRENT APPLICATION NUMBER: US/09/248,796A ; PRIOR FILING DATE: 1999-02-12 ; PRIOR APPLICATION NUMBER: US 60/074,725 ; PRIOR FILING DATE: 1998-02-13 ; PRIOR APPLICATION NUMBER: US 60/096,409 ; PRIOR FILING DATE: 1998-08-13 ; NUMBER OF SEQ ID NOS: 28208 ; SEQ ID NO 20613 ; LENGTH: 722 ; TYPE: PRT ; ORGANISM: Candida albicans ; US-09-248-796A-20613				
 Query Match 7.4%; Score 112.5; DB 4; Length 722; Best Local Similarity 22.2%; Pred. No. 0.031; Matches 58; Conservative 40; Mismatches 90; Indels 73; Gaps 10;				
Qy	38	KTPDETIELSRPKQYSQSEASVLVGDIKVLMDSQKYFEATQ-----	TYYEWCGVVTQ	92
Dd	218	KTLKSTREELNGSKTEILR-----	LKALLRESEDELYQVKQENYKTSVHDYEQDLAQ	269
Qy	93	L-----	LSAYILLDFEYNEKKASAQDKILI--RIILDGGVKKLNEAQS	133
Dd	270	LKVHETLLSRNKDINESLEYKKRSDYYKKULEAESAI	SKRHEEQATKWESRSQ	329
Qy	134	LITSSQFNASSGKLIALDSQLTNDFSEKSYFSQSOVDRIKRAYAGAAAAGIVAGPFGLI	193	
Dd	330	LLLAREE-----	LRTQILIKDPRIKENLEATIEE-----	360
Qy	194	ISVSIAAGVIEGKLIPE-LNNRLKTQN--	FFTLSATVQANKNDIAAKLKATETAAI	250
Dd	361	KNHQLDANKEEIQIQOKLNLYHKNFENKELNKLKEEIKNLNRDL	----FKTDI---	412
Qy	251	GEIKTETETTRFYVDYDDLML	271	
Dd	413	-ETKLKENKLUDDYDVLL	432	
 RESULT 6 US-09-107-532A-5150 ; Sequence 5150, Application US/09107532A ; Patent No. 6583275 ; GENERAL INFORMATION: ; APPLICANT: Lynn A Doucette-Stamm and David Bush ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUT ; NUMBER OF SEQUENCES: 7310 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION ; STREET: 100 Beaver Street				

```

; CITY: Walham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariandelio, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1231
; SEQUENCE DESCRIPTION: SEQ ID NO: 5150:
US-09-107-532A-5150

Query Match 7.3%; Score 111; DB 4; Length 1231;
Best Local Similarity 23.3%; Pred. No. 0.098;
Matches 78; Conservative 58; Mismatches 125; Indels 74; Gaps 17;

QY 6 AECIVVKSATETADGALDLYNK-----YLDQVVPKTFDE-----TIKELS 48
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 AEEIVSMGSAQTIEQITDYLNANGRTGFLNIHLXRPFPVETVQKLPHTVKTIAVL 328
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 49 RFKQESQASVLVGDIVKLLMDSQ-----DKYFEATQVYEWCGVVTQLLSAYILLF 101
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 329 RSKEPGAGGEPILL-DVQSAIYDSELRPAVIGRGYGLGSK-----DVTPDQISA 378
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 DEYNEKASAKOILIRILDGKVLKNEAKSL-LTSSQSFN-----NASGKL 148
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 DELKDPSTRKRKFTIGIVDDVYQSLPEKESLDLTPQTFOAKFWGFGSDGTVCANKSA 438
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 LALDSQLTNDFSKSSYFQSQ-----VDRIR-----KEAVAGAAAGIVA--GPFGLII 194
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 439 IKIIGDHTDKYAQQGYFYDYSKSGGLTVSHLRFGDTPIRSAVLVEHADIVACHTP-AYILH 497
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 195 SYSIAGVIEGKLIPELNNELKTVQFFTSLSATVKQ--ANKDI--DAAKLKLAETEAA 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 SYDLVRLKPGGTF--LLNTLWSEDEQLETHPLPKLAKRYLAENNRIFYTINAMRLAQEVGL 555
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 IGIKIKETETTRF-----YVDYDDLMLSLKGAAKK 280
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 556 GRRINTAMETAFFKLADIIPFDE-VLPLLKEEALK 589
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 7.2%; Score 108.5; DB 4; Length 1066;
Best Local Similarity 21.6%; Pred. No. 0.14;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;

QY 24 LDLYNKYLDQVVPKTFDETIKELSRFKQESQASVLVGDIVKLLMDSQDKYPEATQTV 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 LELYNEELCDLL---STDDTTK--IRIFDDSTKGSVLIQGLEIPIVHSDVDVYKLEKG 213
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 YEWCGVVTQLLSAYILLFDEYNEKASAKOILIRILDGKVLKNEAKSLTSS 138
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 KERRKTATILMNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNVLVDLAGS 265
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 139 QSFNNASGK-----LLALDSQLTN--DFSEKSSYFQSOVDRIKKEAYAGAA 182
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 ENVKAGNEKIRVRETVMNQSLTLGRVITALVDRAPHVYRESKLTLLQESLGRT 325
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 -----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNNRL--KTQNFFTLSATVK 230
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 KTSIIATISPGHKDIEETLSTLEYAHRAKNIQNK--PEVNOQLTKTKVLKEYTE---EID 380
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 QANKDIDAAKLKLAETIA--AIGEIKTETETTRFVYVDLMLSLK 275
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 KLRDLMAARDKNGIYLAETVGEITLKESQNRNELNRMWLLKALK 427
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 7.2%; Score 108.5; DB 4; Length 1066;
Best Local Similarity 21.6%; Pred. No. 0.14;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;

QY 24 LDLYNKYLDQVVPKTFDETIKELSRFKQESQASVLVGDIVKLLMDSQDKYPEATQTV 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 LELYNEELCDLL---STDDTTK--IRIFDDSTKGSVLIQGLEIPIVHSDVDVYKLEKG 213
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 YEWCGVVTQLLSAYILLFDEYNEKASAKOILIRILDGKVLKNEAKSLTSS 138
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 KERRKTATILMNA-----QSSRSHTVFSIVVHIRENGIEGEDMLKIGKLNVLVDLAGS 265
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Qy	139	QSPFNASGK-----LJALDSOLTN--DFSEKSSYFQSOVDRIKREAVAGAA	182
Db	266	ENVSKAGNKGTRVRETWNINQSLTTLGERTVITLVDRAPHPVRESKLTLLQESLGGRT	325
Qy	183	-----AGIVAGPFL-----IISYSIAAGVTEGKILPELANRL--KTVQNFFTSLSATVK	230
Db	326	KTSIIATSPGHKEETLTSLVHAHRAKIQNK--PEVNQKTKTKTLVKEYTE---EID	380
Qy	231	QANKDIDAAKLKATEIA--AIGSIKTTETTRFYVDYDILMLSLK	275
Db	381	KLKRDLMARDKNGYVGLAEETVGYETTLKLESONRELKMLLKALK	427

```

RESULT 9
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

```

Query Match	7.2%;	Score 108.5;	DB 4;	Length 1066;
Best Local Similarity	21.6%;	Pred. No. 0.14;		
Matches 62;	Conservative 62;	Mismatches 110;	Indels 53;	Gaps 12;
Qy	24	LDLYNKYLDQVWPWKTPEDETIKELSRFQKQSYQSEASVLVDGIKVLMLDSQDKFYEAQTQV	83	
Db	159	LEYANEELCDLL---STDDTTK--IRIFDSDTKSGSVIIQGLEBIPVHSXDDVYKLEKG	213	
Qy	84	YEWGVVVTQLLSAYILLFDEYNEKKAQAQDILIRILDDGVK-----KLENAQKSLTSS	138	
Db	214	KERRKTATTLWNA-----QSRSHTVFSVVHIRENGIEGEDMLUKIKLINLVDLAGS	265	
Qy	139	QSFNNASGK-----LLALDLSQLTN--DFSEKSSYPQSQVDRIKREAYAGAA	182	
Db	266	ENVSKAGNEKGI RVETWNIQSLLTLGRVITALVDRAHPVPYRESKLTLLQSSLGRT	325	
Qy	183	-----AGIVAGPGL-----IISYSAAGVIEGKLIPELNNRL--KTVQHFPSLSATVK	230	
Db	326	KTSIIATISPGHKDIBETLSTLEYAHRAKQIQNK--PEVNQKTKTKVLXEYTE---EID	380	
Qy	231	QANKDIDAACKLATEIA--AIGIKTETETTFYVDYDDLMLSLLK	275	
Db	381	KLKRDLMWAARDKNGIYLAEEYTGITUKLSONRELNEKMLLKALK	427	

RESULT 10
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: EPIMERIDIS FOR D
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,

```

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRM
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match          7.2%; Score 108.5; DB 3; Length 3696;
Best Local Similarity 21.5%; Pred. No. 0.95;
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;

QY 11 EVVKSATETADGALDLYNRYKLDQVDPWKTFDETIKLSR-----FKQEYSQ-EA 58
DB 568 EQNKDILIPSNYTLASNYK-----NKLKERAQTVLDEETNNTFNQRYSTQI 616

QY 59 SVLVGDIKVLMM-----DSQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
DB 617 DDLHELQTLTILNRYVSASREINDKAQEMTDVAVYDSTELATTEEKDT--LVQDIENHKNEI 673

QY 112 QKDILIRLDDGVKKLINEAQKSLTTS-----QSFNASKGLLALDLSQLTNDFS 160
DB 674 SNNIDBELTDQGVRYKEAGLHPTLESDPHPVTKPNARQVNNRA-----DQCKTLIRN 727

QY 161 EKSSYFQSQVDRIKRY-EAYAGAAGVAGFPGLIISVSIAGVIEGKLIPELNNRLKTQV 219
DB 728 NHEATTEEQNEAIRQVEAHSSDAIAGE-----AETDTTVNEAR 767

QY 220 NFFTSLSAT-VKQANKDIDAAKCLKLATEIAAIGIKETETETTRFYVDYDMLSLKGA 278
DB 768 DNGTKLIATDVPNPTKAEA---RAAVTNSANSKIDINNNTQATLDERNDALVNRSK 824

QY 279 KKWINTCNEYQ-----QRHGKKTILFEVP 301
DB 825 DEAIQINNTAQGNDDVTEAQNNGTNTIQVP 855

```

```

RESULT 11
US-09-134-001C-4820
, Sequence 4820, Application US/09134001C
, Patent No. 6380370
, GENERAL INFORMATION:
, APPLICANT: Lynn Doucette-Stamm et al
, TITLE OF INVENTION: NUCLEIC ACID AND AMINO
, TITLE OF INVENTION: EPIDERMIDIS FOR DIAGN
, FILE REFERENCE: GTC-007
, CURRENT APPLICATION NUMBER: US/09/134,001C
, CURRENT FILING DATE: 1998-08-13
, PRIOR APPLICATION NUMBER: US 60/064,964
, PRIOR FILING DATE: 1997-11-08
, PRIOR APPLICATION NUMBER: US 60/055,779
, PRIOR FILING DATE: 1997-08-14
, NUMBER OF SEQ ID NOS: 5674
, SEQ ID NO 4820
, LENGTH: 1211
, TYPE: PRT
, ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

```

```

Query Match      7.1%; Score 107; DB 3; Length 1211;
Best Local Similarity 18.3%; Pred. No. 0.24;
Matches 61; Conservative 59; Mismatches 131; Indels 82; Gaps 9;

Qy      7 EQTVEVVK-----SAIETADGALDLYNKYLDQVWPKTFDETIKELSRFKQEYSQASV 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      185 ESSAGVLKYKRAESTQKLDHTEDNLRVVEDILYDLGRVFLPEEAAATAKEVKKLSKE 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      61 LVGDIKVLMDSQDYFEATQTVYEMC-----GVVTQLLSAYILLFDSEYN 106
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 245 MEQSDVIVTVSDIDHYTDNQRDLERLNLKSOAEKEGQQAQINOLLOKY----- 295
Qy 107 KKAQAQDIIIRILDGKVLNEAQSLLTSSQSFNNASGKLIALDLSQLTND-----FS 160
Db 296 -KGRQON-----DYDIEKLN---YELVKATENYEQLSGKLVLEERKKNQSETNARYE 345
Qy 161 EKSSYFQSDVRIRKEAYAGAAAGIVAGPPGLIISYIAAGVIEGKLIPELNNRL-----K 216
Db 346 EELDNLESQDSIKNEKAQNE-----KLLADLKNKQKQLNK 381
Qy 217 TVQNFTSLSATVQKANKDIDAAKKLATEIAAIGELKTETETTRFYVDYDIDLMLSLKKG 276
Db 382 EVOELSLLYISDQHDKELEIKNSYITLMSQSVVNDIRPLEHTINENEAKSRDLS 441
Qy 277 AAKMINTCNEYQQ-----RHGKKTLEFV 300
Db 442 RLVEAFNQLKDIIQNIQTQKEYQSSKSKMEKV 474

RESULT 12
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-08-592-126-148

Query Match 7.0%; Score 106.5; DB 2; Length 1312;
Best Local Similarity 18.9%; Pred. No. 0.3;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
Qy 10 VEVKSAIEFADGALDLYNKVLDQVLPKTFDETIK-----ELSRFQKEY 54
Db 502 MEVLSQNEKAD--LDRTLRKLDQEMQLNHHYTRTQMEMLTKDKADKDEQIKIKSRH 559
Qy 55 SQEASVLGDIKVLIMSQDKFYFATQTVYVCGVWVQLLSAYILLDFEYNEKKAQKD 114
Db 560 SDELTSLIG-----YFPNKKQLEDWLHSSK-----EINQTR----- 591

Qy 115 ILIRILDGKVLNEAQSLLTSSQSFNNASGKLIALDLSQLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKHNINELKREEQQLSS-YEDKLFVCGSQDFES 640
Qy 169 QVDRIRKE-----AYAGAAAGIVAGPP----- 190
Db 641 DLRLKEIEKSKORAMLAGATA--VYSOFITQLTDENOSCCPVORVPQTEAELOQEV 698
Qy 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217
Db 699 SLOSKLRLADPKLKSTESLKKKEKRDDEMGLVPMRQSIID--LKEKEIPELNNKLON 756
Qy 218 VQNFTSLSATVQKANKDIDAAKKLATEIAAIGELKTETETTT-----RFVVDY 266
Db 757 V-----NRDIQRLKNDIEBEQETLLGTIMPESAKVCLTDTVIMERFOMEL 802
Qy 267 DDLMLSLKGAQK-----KMINTCNEYQORHGKK 295
Db 803 KDVERKIAQAQAKLQGLDLDRTVQOVNQEKQKH 838

RESULT 13
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cdNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 7.0%; Score 106.5; DB 2; Length 1312;
Best Local Similarity 18.9%; Pred. No. 0.3;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;


```

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18798

Query Match          7.0%; Score 105.5; DB 4; Length 924;
Best Local Similarity 17.8%; Pred.No. 0.22; Indels 77; Gaps 8;
Matches 51; Conservative 58; Mismatches 100;

QY    26 LYNKYLDQVIPWKPTDETIKELR-----FKOEYSQEASVLVGDIKULLMDSQDKVFATQ 81
DB    233 LHKKYVD-----SIKDLKDFLAFKOFEAGIINVL-NLDLAQLSELKQKOIDLLEN 281

QY    82 TVYEWCGVVQTLLSAYILLFDEVNEKSAQAOKDILIRILDGKVKKLINEAQSLTSSQS F 141
DB    282 TKND--QIKELVSEHELOIEKISKDLTEKFKLIVETQLLSKHESHVQQFTTKELIAES EK 338

QY    142 NNA-----SGKIHALDSQLTNDFSSEKSYFSQSOVDRIKEAYAGAAGIVA 187
DB    339 QOVEEBELENLKTHAKDSARILELETQLSDAAEKSES DYKLT-----382

QY    188 GPGLLIISVIAGVIEGKLIPELANRLKTGVNFFTSLSATVKQANKOIDDAAKLKLAT E I 247
DB    383 -----TSEIVNLDKSQITFLKANLNK-----EEEREIQNKKLQDVSEL 421

QY    248 AAIGEIKETEPTTRFYVDYDDMLMLLKGAAXMINTCNEYQRHG 293
DB    422 KELKELKVE-----ELSNNLLKQLQBELHRKEIELNQLEKJLHG 459


RESULT 18
US-09-538-092-807
; Sequence 807, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformatter Version 0.9
; SEQ ID NO 807
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YPR070W
US-09-538-092-807

Query Match          6.9%; Score 104.5; DB 4; Length 566;
Best Local Similarity 25.4%; Pred.No. 0.13;
Matches 57; Conservative 29; Mismatches 91; Indels 47; Gaps 11

QY    15 SALETADGALDYNNK-----LDQVIPWKPTDETI-KELSRFKQ-----52
DB    6 SYVETLDSMLELFKDYPGSIVLENTLCQTIGLESFTELSENLSRLSTASKIIVID V 65

QY    53 EYSQEASVLVGDIKULLMDSQB--YFEATQTVYEWCGVVQTLLSAYILLFDEVNEK KAS 110
DB    66 DYNKQD-RIQDWKLVLAGNFDFYFNQRDGEHKEKNILLSLTYPDLKAFAHNILK EL 124

QY    111 AQKDILIRILDGKVKKLN-EAQSKILTSSSQSFNNAAGKLLALDSQTLTNDFSSEK S Y --- 165

```

```

Db      125 YLDAYSHIESDSTSHNGSSDKLSNNAFNN-QGKL-----DLKYFTTELSHYIROC 178
QY      166 FQSQVDRIRKEAYAGAAAGIVAGPFGLIISYSAAGVIEGKLIP 209
        ||      :      ||      ||      ||      ||      ||      ||      ||
Db      179 FQDNCCDFKVRTNLNDKFGI-----YILTQG-INGKEVP 211

RESULT 19
US-09-107-532A-3944
; Sequence 3944, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                               ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3944:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1196
; SEQUENCE DESCRIPTION: SEQ ID NO: 3944:
US-09-107-532A-3944

```

```

Query Match      6.9%; Score 104.5; DB 4; Length 1196;
Best Local Similarity 19.9%; Pred. No. 0.41;
Matches 58; Conservative 45; Mismatches 92; Indels 97; Gaps 10;

QY      7  EQTVEVVK-----SAIETADG---ALDLYNKYLDQVTPKWTPTDTIKELSRKQE 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      165  EETAGVLVLYKQKKKAAEQKUFETEDNLSRVODILHELEEQTLPLAAOSEAAKEFLURKET 224

QY      54  YSQ--EASVLVGDIIKVLAMDSDQKYFEATQTVYEWGVVTVLLSAYILLFDYENKCSAKAQ 112
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      225  LTQTDVSLVMAEIKTKKD-----WDNKQAQLA 252

QY      113  KQILIRILDGVKKLNEA---QKSLTTSQSFNNAGSKLLALDSQLTNDPFSEKSYFSQSQ 169

Db      253  KFNL-----ELGKLSIESIOEISILAKORENACADRLIEKNQOVLILDSEKLLKOTEGO 306

```

```

Qy 170 VDRTRKAYAGAAAGIVAGPGLIISYSIAAGVIEGKULPELNRLKTVQNFFTLSATV 229
Db 307 KVVQERT-----KHTQKSSQEYQTSLAEEAQ 332
Qy 230 KOAN--KDIIDAAKLKLA---TEI--AAIGEIKTETETTRFVVDYDDLMLSL 273
Db 333 KKVGHPEKIQESLMKAAAEKETEIQKAEANLIKQOELEKYQSTKELLAEL 384

RESULT 20
US-09-540-236-2753
; Sequence 2753, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXY
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2753
; LENGTH: 718
; TYPE: PRP
; ORGANISM: M.catarrhalis
US-09-540-236-2753

```

Query Match	6.8%; Score 103.5; DB 4; Length 718;
Best Local Similarity	20.0%; Pred. No. 0.24;
Matches	66; Conservative 57; Mismatches 138; Indels 69; Gaps 12;
Qy	6 AQTVEVVKSAIETADGALDLYNKYLDQVTPWKTFDETIKELSRFKQE-----VSQE 57 : : : : : : : : : : Db 103 ATEKLEVAKEATQ-----DKBKTOSLVEDIKKAQSLOEQEDAADV EALKQA 149 : : : : : : : : : : Qy 58 ASVLVGDIKV---LLMDSQDKYFEATQTVYEW C--GVVTQLLSAYILLFDBYN-----E 106 : : : : : : : : : : Db 150 ASDKVETTKEAQSLKDDATQTPESAKQAVEGKVEAIKEQVLQDVDSLKD TTDODNTDQD 209 : : : : : : : : : : Qy 107 KGAASAQDILIRLDGCVKLNEBAQSKLLTSSQS F--NNAGKL-----LALD--S QUTND 158 : : : : : : : : : : Db 210 QEKQTLKDKAVQAATAAKRVEDVDVDVKHTTES FKNFTASGKIDEIKQA AVDTKEEVKSQ 269 : : : : : : : : : : Qy 159 PSEKSSYFSQOVDRIRKEA VAGAAGIVAGPFGL IISYSIAANGVIECKLP ELNRLKTV 218 : : : : : : : : : : Db 270 LSQLKADALKSGSELKGTQTAANDAITEAQ--AAVWSGSVAAA-----DSAQSTA 318 : : : : : : : : : : Qy 219 QNFPTSLSATVQANKIDDAKLKLAPEIAAIGBIKT ETETRTFPYVDYDDLMLSLKGAA 278 : : : : : : : : : : Db 319 QSAKDKLNQLPFOEQKSLDELKVOELGEKFQGT EKINAVSNV-----DLATQVIKEEA 371 : : : : : : : : : : Qy 279 KKMINTC-----NEYQQRHGKKTL 297 : : : : : : : : : : Db 372 QALOTNAOESLOAKAAGEYDYATHDPKGL 401 : : : : : : : : : :

RESULT 21
US-08-477-831C-2
; Sequence 2, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUWEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA: US/08/477,831C
; APPLICATION NUMBER: US/08/477,831C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PIERRE, MARGARET A.
; REGISTRATION NUMBER: 30,709
; REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: /desc = "RHAMM I protein"
;
; US-08-477-831C-2
;
; Query Match 6.8%; Score 103; DB 4; Length 606;
; Best Local Similarity 19.8%; Pred. No. 0.2;
; Matches 66; Conservative 69; Mismatches 131; Indels 68; Gaps 17;
;
; QY 1 MTSIFAQTVVEVKSIAETADGALDLYNKYLDOVIPWKTFTDTIKELSRFKQYEQSASV 60
; DB 228 LNLRLREKEVELEKHIAHAQAIIAQEKYNTA---QSLRDVTAQLESVQEKYNTAQS 284
; QY 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKAQAOKDILIRIL 120
; DB 285 L-RDVTQALESQEKYNTAQSRLD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 334
; QY 121 DGCVKLNKAQSL---LTSSQSFNNASGKLALDSQLTN-DFSEKSYFQSOVDRIKKE 176
; DB 335 ESQVEKYNTAQSRLDVTQALESYSKSTLKEIE-DLKENLTLOEKVMAEKSVEDVQQQ 393
; QY 177 AVAGAAAGIVAGPFGILISYSIAAGVIEGKLIPELNNR-----LKTQVQNF-----TSL 225
; DB 394 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETITSSFLEKITDL 436
; QY 226 SATVQANKDIDAANKLATEIAAIGETKE--TETTRFYVDYDDL----- 269
; DB 437 KNQLRQODEDFRQLEEKGRKTAENVMTELTMTWINKWRLLYDELYEYKTKPFQQLDAF 496
; QY 270 ---MLSLK--GAACKMINTCNE-YQORHGKKTIL 297
; DB 497 EAEKQALLNEHGATQEQNLKIRDSYAQLLGHQNL 530

```

```

;
; RESULT 22
; US-08-477-831C-11
; Sequence 11, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA: US/08/477,831C
; APPLICATION NUMBER: US/08/477,831C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PIERRE, MARGARET A.
; REGISTRATION NUMBER: 30,709
; REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: /desc = "RHAMM I-2a"
;
; US-08-477-831C-11
;
; Query Match 6.8%; Score 103; DB 4; Length 631;
; Best Local Similarity 19.8%; Pred. No. 0.22;
; Matches 66; Conservative 69; Mismatches 131; Indels 68; Gaps 17;
;
; QY 1 MTSIFAQTVVEVKSIAETADGALDLYNKYLDOVIPWKTFTDTIKELSRFKQYEQSASV 60
; DB 253 LNLRLREKEVELEKHIAHAQAIIAQEKYNTA---QSLRDVTAQLESVQEKYNTAQS 309
; QY 61 LVGDIKVLMLDSQDKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKAQAOKDILIRIL 120
; DB 310 L-RDVTQALESQEKYNTAQSRLD---VTAQLESEQ-----EKYND-TAQLSLRDVTAQ-L 359
; QY 121 DGCVKLNKAQSL---LTSSQSFNNASGKLALDSQLTN-DFSEKSYFQSOVDRIKKE 176
; DB 360 ESQVEKYNTAQSRLDVTQALESYSKSTLKEIE-DLKENLTLOEKVMAEKSVEDVQQQ 418
; QY 177 AVAGAAAGIVAGPFGILISYSIAAGVIEGKLIPELNNR-----LKTQVQNF-----TSL 225
; DB 419 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETITSSFLEKITDL 461
; QY 226 SATVQANKDIDAANKLATEIAAIGETKE--TETTRFYVDYDDL----- 269
; DB 462 KNQLRQODEDFRQLEEKGRKTAENVMTELTMTWINKWRLLYDELYEYKTKPFQQLDAF 521
; QY 270 ---MLSLK--GAACKMINTCNE-YQORHGKKTIL 297
; DB 522 EAEKQALLNEHGATQEQNLKIRDSYAQLLGHQNL 555

```

```

;
; RESULT 23
; US-09-538-092-1078
; Sequence 1078, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 1078

```



```
; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Query Match          6.7%; Score 102; DB 4; Length 1976;
Best Local Similarity 21.1%; Pred. No. 1.6;
Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;

QY 52 QEYSQASVLVGDIKVLLMSQDKYFEATQTVYEWGVVTVQLLSAYILLDFDEYNEKASA 111
Db 972 EKVTAEAKIKMEEIILLDDQSKTKKEKLM-----DRIAECSQSLAESEKAKNLA 1026
QY 112 ----QXDLIRLLDDGVKLLNEAQKSLTSSQSFNNASGKLLALDSQLTNDFSEKSSYFQ 167
Db 1027 KIRKQEVMSLDLEERLKKEKTRQBLEKARK-----LDGE-TTDLQDQIAELQ 1075
QY 168 SQVDRI-----RKEAYAGAAAGIVAGPFGLLIISYSIAAGVIEG-----KLIPELNRR 214
Db 1076 AQIDELKQLAKKEEELQGA-----LARGDDETLHKNNALKVVRLELQ 1119
QY 215 LKTVQNPFTSLATVQKANKDIDAAKLLATEIAAIGETETTRFYVDYDDMLSL 274
Db 1120 IABLQEDFESEKASRNKAEKQ-----KRLSELEA-----LKTELEDT-----LD 1160
QY 275 KGAAKWMINTCNEYQQRHGKKTIFE 299
Db 1161 TTAQQELRTKREQEVAELKALEE 1185

RESULT 24
US-08-402-217A-3
; Sequence 3, Application US/08402217A
; Patent No. 5587301
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,217A
; FILING DATE: 10-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: 33954
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Query Match          6.7%; Score 101; DB 1; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.22;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;

QY 1 MTSIPAEQVEVVVKSIAETADGALDLYNKYLDQVIPWKTFFDETIKLSRFKQYSQEASV 60
Db 99 LONLLREKEVELEKHTAAHAQAAILIAQEKYNTA---QSLRDVTAQLESQEKYNDTAQS 155
QY 61 LVGDIKVLMSQDKYFEATQTVYEWGVVTVQLLSAYILLDFDEYNEKASAQKDLIRIL 120
Db 156 L-RDVTVAQLESQEKYNDTAQSLRD---VTAQLESQ---EKYND-TAQLSLRDVTAQ-L 205
QY 121 DGQVKLNEAQKSLTSS---QSFNNASGKLLALDSQLTN-DFSEKSSYFQSOVDRIKE 176
Db 206 ESQVEKYNDTAQSLRDVSAQLESYKSTLKEIB-DLKJENLTLQEKVAMAEKSVEDVQQ 264
QY 177 AVAGAAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNR-----LKTQVONFF---TSL 225
Db 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETSSFEKITDL 307
QY 226 SATVQKANKDI-----DAAKLKLATEIA-----AIGETETTRFYVD-Y 266
Db 308 KNQLRQODEDFRKQLEBKGRKTAENVMTELTMEINKWRLLYEELVEKTKPFQOQLDAF 367
QY 267 DDLMSLLK--GAAKWMINTCNE-YOORHGKKTIL 297
Db 368 EAEKQALLNEHGATQEQQLNKIRDSYAQLLGHQNL 401

RESULT 25
US-08-700-178-3
; Sequence 3, Application US/08700178
; Patent No. 5783669
; Patent No. 5783669 5700912
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,178
; FILING DATE: August 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
```

```

; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; LIBRARY: mouse
; CLONE: GI 53979
US-08-700-178-3

Query Match          6.7%; Score 101; DB 1; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.22;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;

QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQYSQASV 60
DB 99 LDNLLREKEVELEKHIAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKVLMSQDKYFEATQTVYEGCVVTTQLLSAYILLFDEYNEKASAKDILIRIL 120
DB 156 L-RVTAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSRDVTAQ-L 205
QY 121 DGVKKLNEAOKSLTSS---QSFNAGSKLLALDSQLTN-DFSEKSSYFQSOVDRIKE 176
DB 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLENLTQEKVMAEKSVEDVQQQ 264
QY 177 AVAGAAAGIVAGPFLIISYSIAAGVIRGKLIPELNNR-----LKTQVQNF---TSL 225
DB 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEIEIKETSSFLEKITDL 307
QY 226 SATVQKANKDI-----DAAKKLATEIA-----AIGIKETETETTFYVD-Y 266
DB 308 KNQLRQODEDFRKQLEEKGRKTAENVMTELMEINKWRLLYBELYEKTKPFQQLDAF 367

RESULT 26
US-08-995-654-3
; Sequence 3, Application US/08995654
; Patent No. 6025138
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig
; APPLICANT: Seilhamer, Jeffrey
; TITLE OF INVENTION: HVALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,654
; FILING DATE: December 22, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/700,178
; FILING DATE: August 20, 1996
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0028-2 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-995-654-3

Query Match          6.7%; Score 101; DB 3; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.22; 127; Indels 68; Gaps 18;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;

QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQYSQASV 60
DB 99 LDNLLREKEVELEKHIAHAQAIIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKVLMSQDKYFEATQTVYEGCVVTTQLLSAYILLFDEYNEKASAKDILIRIL 120
DB 156 L-RVTAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSRDVTAQ-L 205
QY 121 DGVKKLNEAOKSLTSS---QSFNAGSKLLALDSQLTN-DFSEKSSYFQSOVDRIKE 176
DB 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLENLTQEKVMAEKSVEDVQQQ 264
QY 177 AVAGAAAGIVAGPFLIISYSIAAGVIRGKLIPELNNR-----LKTQVQNF---TSL 225
DB 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEIEIKETSSFLEKITDL 307
QY 226 SATVQKANKDI-----DAAKKLATEIA-----AIGIKETETETTFYVD-Y 266
DB 308 KNQLRQODEDFRKQLEEKGRKTAENVMTELMEINKWRLLYBELYEKTKPFQQLDAF 367

RESULT 27
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-310-187A-1

Query Match          6.7%; Score 101; DB 3; Length 1939;
Best Local Similarity 22.5%; Pred. No. 1.9;
Matches 75; Conservative 44; Mismatches 128; Indels 86; Gaps 13;

QY 5 PAEQTVVVKSAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQYSQASV 64

```

Db 1173 FQWRDLLEATLQHEATAAALRKKHADV---AELGEQIDNLRVQKLEKEKS-----E 1225
Qy 65 IKVLLMDSQDKYFEATQTVYVCGVTTQLLSAYILLFDEYNEKASAKQDILIRILDGV 124
Db 1226 FKLELDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258
Qy 125 K-----KLINEAKSL---LTSQSFNNASGKLL-----ALDSQLTNDPSEKSSYFOSQV 170
Db 1259 NEYRVKLEEAQRSINDFTTQRAKLQTEGELARQLEKEKALISQLTRG---KLSYQQOME 1315
Qy 171 DRIRKEAYAGAAAGIVAGPGLIISYISIAAGVIEGKLIPE-----LNRL 215
Db 1316 DLKQLEEGKAKNALA-----HALQSAHDCDLLREYBEETEAKAELQVLSKAN 1367
Qy 216 KTVQNFTSLSATVKQANKDIDAAKLKATE-----IAAIGEIKTETETT---RFVVDY 266
Db 1368 SEVAQWRTKYETDAIQRTTELEEAKKKLAQRLQDAEBAEAVNAKCSLSLEKTKHRLQNEI 1427
Qy 267 DDLMLSLKGAACKMINTCNEYQORHGKKTLPF 299
Db 1428 EDLMVDVRSNAAA---AALDKQKRNFDKILAE 1457
RESULT 28
US-09-538-092-917
; Sequence 917, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 917
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P13533
US-09-538-092-917

Query Match 6.7%; Score 101; DB 4; Length 1939;
Best Local Similarity 22.5%; Pred. No. 1.9;
Matches 75; Conservative 44; Mismatches 128; Indels 86; Gaps 13;
Qy 5 FAEQTVVVKSAIETADGALDLYNKYLDQVLPKTFDETIKELSRFKQEYSQEAASVLGD 64
Db 1173 FQWRDLLEATLQHEATAAALRKKHADV---AELGEQIDNLRVQKLEKEKS-----E 1225
Qy 65 IKVLLMDSQDKYFEATQTVYVCGVTTQLLSAYILLFDEYNEKASAKQDILIRILDGV 124
Db 1226 FKLELDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258
Qy 125 K-----KLINEAKSL---LTSQSFNNASGKLL-----ALDSQLTNDPSEKSSYFOSQV 170
Db 1259 NEYRVKLEEAQRSINDFTTQRAKLQTEGELARQLEKEKALISQLTRG---KLSYQQOME 1315
Qy 171 DRIRKEAYAGAAAGIVAGPGLIISYISIAAGVIEGKLIPE-----LNRL 215
Db 1316 DLKQLEEGKAKNALA-----HALQSAHDCDLLREYBEETEAKAELQVLSKAN 1367
Qy 216 KTVQNFTSLSATVKQANKDIDAAKLKATE-----IAAIGEIKTETETT---RFVVDY 266
Db 1368 SEVAQWRTKYETDAIQRTTELEEAKKKLAQRLQDAEBAEAVNAKCSLSLEKTKHRLQNEI 1427

Qy 267 DDLMLSLKGAACKMINTCNEYQORHGKKTLPF 299
Db 1428 EDLMVDVRSNAAA---AALDKQKRNFDKILAE 1457
RESULT 29
US-09-538-092-844
; Sequence 844, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 844
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P04264
US-09-538-092-844

Query Match 6.6%; Score 100.5; DB 4; Length 643;
Best Local Similarity 22.2%; Pred. No. 0.4;
Matches 72; Conservative 58; Mismatches 124; Indels 71; Gaps 15;
Qy 5 FAEQTVVVKSAIETADGALDLYNKYLDQVLPKTFDETIKELSRFKQEYSQEAASVL 61
Db 199 FLEQGNQVLQTKWELLQQVDTSTRTHN-----LEPY---FESFINLRRRVYDQLKSDQSL 251
Qy 62 VGDIK---VLLMDSQDKYPE---ATQTVYVCGVTTQLLSAYILLFDEYNEKASAKQD 114
Db 252 DSELKQMDVEDYRNKYDEINKRTNAENEFYTIKVDGAYMTKVD-LQAKLDNLQOE 310
Qy 115 ILIRILDGVYKLENEAKSLTSSQSFNNA-----SKLLALDSQL-----TNDFSEKS 163
Db 311 I-----DPLTALYQAELSQMQTQISETNVILSMNDRSLDLSIIAEVKAQNEEDIAQS 364
Qy 164 -----SYFQSQVDRIKREYAGAAAGIVAGPFG-----LIISYISIAAGVIEGKLIPELNN 213
Db 365 KAAEASLYQSKYBELQ-----ITAGRHGDSVRNSKIEISELNRVIQ-RLRSEIDN 413
Qy 214 RLKTVQNFTSLSATVKQANKDIDAAKLKATEIAAIGEIKTETETTFRFVYVDDLM--- 270
Db 414 VKQISNLQOOSISDASQGENALKDKNLDLEALQQAQK--EDLARLLRDTQELMNTK 471
Qy 271 -----LSLLKGAACKMINTC 285
Db 472 LALDLEIATVYRTLLEGESRMSGEC 496

RESULT 30
US-09-328-352-5793
; Sequence 5793, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5793
; LENGTH: 1454
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5793

Query Match          6.6%; Score 100.5; DB 4; Length 1454;
Best Local Similarity 19.6%; Pred. No. 1.4;
Matches 80; Conservative 57; Mismatches 133; Indels 139; Gaps 18;

QY      11 EVVKAIEET-----ADGALDLYNKYLDOVIPWKTFDETIKE-----LSRF 50
DB      278 QVLGAGSVQTLASNLDLIADGALVGVIGYITRAILMKS--AAIKEGHASTLASRQASVLNA 335
QY      51 KOEYSOEASVL-----VDGIKVLIMDSODKYFEAT-----QTVYE 85
DB      336 QAIEAERTAAALNAAKHLANVRATNAETQAK-FGATAATRYAQAAVAATTAATNAQTAAG 394
QY      86 -----WCGVVTTQLLSAYILLFPEYNEKKASAOKDI-----LI 117
DB      395 IKLNTATSIAGRLAGAFGLGGWAGVATLGVMGLAAAYSFFNNKAEERAKLAEOAKVA 454
QY      118 RIILDDGVKKL--NEAQSKLLTSSQSFFNNSAGKLLALDSQLTNDF---SEKSSSYFSQSOVDR 172
DB      455 EKADEELKKLTGNDAKAVNDLITAP-NAQNKALESRRVGSALIDIENVARGNREVEK 513
QY      173 IRKEAYAGAAGIVAGPFGGLIISYSIAGVIEGKLIPELNNRLKTQVFNFSTLSATVKQ- 231
DB      514 ISQEARTGT-----ISYTEAI-----ERLNKIPLTDLYNLNKKQAQY 552
QY      232 ----ANKDIDAACKL-----ATSIAGIETETE-TTRFYVDY- 266
DB      553 DDNASKASLSAEKLLRVEVKLGNGEAQNAIQHQKDADALGNTATEAEKATKALQDYQ 612
QY      267 -----DDLMLS--LLKGAAKKMTINTCNEYQQRHGKTKTLFEVPDVAS 305
DB      613 AKQKSDVIDSIYKSLMGDKTYTVAQANAILELQKAGMSAILSKXDEIDS 661

RESULT 31
US-09-538-092-1077
; Sequence 1077, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormat Version 0.9
; SEQ ID NO 1077
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Query Match          6.6%; Score 100; DB 4; Length 1960;
Best Local Similarity 19.1%; Pred. No. 2.5;
Matches 58; Conservative 46; Mismatches 120; Indels 80; Gaps 9;

QY      7 EQTVFVVKSAIE----TAGALDLNKNYLDVIPWKTFDETIKE-----LSRFKQEYS 55

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5793
; LENGTH: 1454
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5793

Query Match          6.6%; Score 100.5; DB 4; Length 1454;
Best Local Similarity 19.6%; Pred. No. 1.4;
Matches 80; Conservative 57; Mismatches 133; Indels 139; Gaps 18;

QY      11 EVVKAIEET-----ADGALDLYNKYLDOVIPWKTFDETIKE-----LSRF 50
DB      278 QVLGAGSVQTLASNLDLIADGALVGVIGYITRAILMKS--AAIKEGHASTLASRQASVLNA 335
QY      51 KOEYSOEASVL-----VDGIKVLIMDSODKYFEAT-----QTVYE 85
DB      336 QAIEAERTAAALNAAKHLANVRATNAETQAK-FGATAATRYAQAAVAATTAATNAQTAAG 394
QY      86 -----WCGVVTTQLLSAYILLFPEYNEKKASAOKDI-----LI 117
DB      395 IKLNTATSIAGRLAGAFGLGGWAGVATLGVMGLAAAYSFFNNKAEERAKLAEOAKVA 454
QY      118 RIILDDGVKKL--NEAQSKLLTSSQSFFNNSAGKLLALDSQLTNDF---SEKSSSYFSQSOVDR 172
DB      455 EKADEELKKLTGNDAKAVNDLITAP-NAQNKALESRRVGSALIDIENVARGNREVEK 513
QY      173 IRKEAYAGAAGIVAGPFGGLIISYSIAGVIEGKLIPELNNRLKTQVFNFSTLSATVKQ- 231
DB      514 ISQEARTGT-----ISYTEAI-----ERLNKIPLTDLYNLNKKQAQY 552
QY      232 ----ANKDIDAACKL-----ATSIAGIETETE-TTRFYVDY- 266
DB      553 DDNASKASLSAEKLLRVEVKLGNGEAQNAIQHQKDADALGNTATEAEKATKALQDYQ 612
QY      267 -----DDLMLS--LLKGAAKKMTINTCNEYQQRHGKTKTLFEVPDVAS 305
DB      613 AKQKSDVIDSIYKSLMGDKTYTVAQANAILELQKAGMSAILSKXDEIDS 661

RESULT 31
US-09-538-092-1077
; Sequence 1077, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormat Version 0.9
; SEQ ID NO 1077
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Query Match          6.6%; Score 100; DB 4; Length 1960;
Best Local Similarity 19.1%; Pred. No. 2.5;
Matches 58; Conservative 46; Mismatches 120; Indels 80; Gaps 9;

QY      7 EQTVFVVKSAIE----TAGALDLNKNYLDVIPWKTFDETIKE-----LSRFKQEYS 55

```


APPLICATION NUMBER: US/08/938,105

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Crook, Wamell M.

REGISTRATION NUMBER: 31,071

REFERENCE/DOCKET NUMBER: 3595-4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1886 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-938-105-3

Query Match

Best Local Similarity 21.3%; Pred. No. 3.7;

Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;

5 PAQTEVYVKSIAETADGALDLYNKYLDQVTPWKTFTETIKELSRFKQYEQEASVLVGD 64

1120 FQKQERDLEATLQHEATAAALRKHADSV---AELGEQIDNLQVVKQLEKES-----E 1172

65 IKVLLMSQDKYFEATQVYEWGCVTQLLSAYILLFDEYNEKKAQAQKDLIRLLDDGV 124

1173 FKLELDD-----VTSHMQII-----KAKANLEKVSRTLEDOA 1205

125 K----KLNEAQSLLTSSQSFNNASGKLLALDSQLTNDSEKS-----SYFSQ 169

1206 NEYRVKLEAQSLL-----NDFTQRAKQLENGELARQLEKEALINQLTGRKLSYTOOM 1261

170 VDIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPE-----LNNR 214

1262 EDLKRQLEBEGKAKNLA-----HALQSAHDCDLREQYEEEMAEKAELOQVLSKA 1313

215 LKTVQNFETSLATVKQANKDIDAARKLATE-----IAAIGIKTETETTT---RFYVD 265

1314 NSEVAGWRTYETDAIORTTELEAKKLAQRLDAAEAVEAVNAKCSLEKTKHRLQNE 1373

266 YDDMLSLKGAAGKAMINTCNEYQOQHGKKTIFE 299

1374 IEDLMVDVRSNAAA---AALDKQKQNFDKILAE 1404

RESULT 38

US-09-252-991A-31794

Sequence 31794, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 31794

LENGTH: 669

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31794

Query Match

Best Local Similarity 21.8%; Pred. No. 0.84;

Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

6.4%; Score 97.5; DB 4; Length 669;

193 VE---SQADGMVAG-----MTITDRKKAQDFSVFVDSGQIA 229

229 VKQAN---KDIDAAK---LKLATEIA-AIGBIKTETETTRFYVDYDDMLSL 274

230 VKGNDKIKSYDLDKGVKIGTESADPLEKXKKYDYSIKYLDTTDALYSAL 284

RESULT 36

US-09-328-352-5599

Sequence 5599, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328.352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 5599

LENGTH: 733

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-5599

Query Match

Best Local Similarity 22.0%; Pred. No. 0.86;

Matches 65; Conservative 43; Mismatches 86; Indels 102; Gaps 14;

12 VVKSIAETADGALDLYNKYLDQVTPWKTFTETIKELS-----RFKQYS-QEASVLVGD 64

201 VMKVAIFTKD-----QLDSKY-----NKKLSIPAAVDNINSSYVAERGLTG- 244

65 IKVLLMSQDKYFEATQVYEWGCVTQLLSAYILLFDEYNEKKAQAQKDLIRLLDDGV 124

245 ILGLNVOGYDKH-----ITQVNLAILVTYGAQNVRRSAESAOTLKFLEQL 292

125 -----KLNEAQSLLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRI- 174

293 PDLKKQLDAAERQFNKFRQYN-----TVDVTKESLYLTQSITLETAKAEL 339

175 KEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTQVNFETSLATVKQANK 234

340 EQKQAEWAAYTA-----EHPAMREINGQITAINKQIGELNSTLKQL-P 382

235 DIDAQKLAETAAIGBIKTETETTRFYVDYDDMLSLKGAAGKAMINTCNEYQO 290

383 DVORQYLYREV-----EVKTQ-----LYTALL-----NSYQQ 411

RESULT 37

US-08-938-105-3

Sequence 3, Application US/08938105

Patent No. 6353151

GENERAL INFORMATION:

APPLICANT: Leinwand, Leslie A.

APPLICANT: Vikstrom, Karen L.

TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESS: Sheridan Ross P.C.

STREET: 1700 Lincoln St., Suite 3500

CITY: Denver

STATE: CO

COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

QY 33 QVTPKTFDETIKELSRFKQY-----SQEASVL-----VGDIKVLLMDSQDK 75
Db 112 EVVPDKQVDDIDAEIAFLKQALGVDRADMRALSKLASQURKEARALFDVILMLDDASI 171
QY 76 YFEATQTVY--EWC-GVVTQLLSAYI-----LLFDEYNEKKASQAQKOILIRILDDGVKKLN 128
Db 172 GNEVKRIIRTGWAQAGALQVVMHVQRFELMDDAYLRRASDVKDIGRELL-----AYLQ 227
QY 129 EAKSLLT-----SQSFNNA-----SGKLLALDSQLTNDPSEKSSYFQSQVDRIRK 175
Db 228 EERKQNLTYPEQTIIVSEELSPAMLGEVPEGRVLGLVSLVSG-----NGSHVAILAR 279
QY 176 EAVAGAAAGVAGPFGGLIISYSIAAGVIEGKLIPELNNRLKTVQNPFSLSATVKQANKD 235
Db 280 AMGIPTVMGAVDLPYSKVDGIDLIVDGYHGEV-----YTNPSAELVRQYSD 325
QY 236 IDAAKLKLAATEIAAIGIKTET 257
Db 326 VVAERELSKGLAALRLPCET 347

RESULT 39

US-09-199-637A-170
; Sequence 170, Application US/09199637A
; Patent No. 6355411

GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Wiklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John

TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

SEQUENCES AND USES THEREOF

; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-170

Query Match 6.4%; Score 97.5; DB 3; Length 759;
Best Local Similarity 21.8%; Pred. No. 1;
Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

QY 33 QVTPKTFDETIKELSRFKQY-----SQEASVL-----VGDIKVLLMDSQDK 75
Db 202 EVVPDKQVDDIDAEIAFLKQALGVDRADMRALSKLASQURKEARALFDVILMLDDASI 261
QY 76 YFEATQTVY--EWC-GVVTQLLSAYI-----LLFDEYNEKKASQAQKOILIRILDDGVKKLN 128
Db 262 GNEVKRIIRTGWAQAGALQVVMHVQRFELMDDAYLRRASDVKDIGRELL-----AYLQ 317
QY 129 EAKSLLT-----SQSFNNA-----SGKLLALDSQLTNDPSEKSSYFQSQVDRIRK 175
Db 318 EERKQNLTYPEQTIIVSEELSPAMLGEVPEGRVLGLVSLVSG-----NGSHVAILAR 369
QY 176 EAVAGAAAGVAGPFGGLIISYSIAAGVIEGKLIPELNNRLKTVQNPFSLSATVKQANKD 235
Db 370 AMGIPTVMGAVDLPYSKVDGIDLIVDGYHGEV-----YTNPSAELVRQYSD 415
QY 236 IDAAKLKLAATEIAAIGIKTET 257
Db 416 VVAERELSKGLAALRLPCET 437

RESULT 40

US-09-248-796A-16013
; Sequence 16013, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16013
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16013

Query Match 6.4%; Score 97.5; DB 4; Length 862;
Best Local Similarity 21.8%; Pred. No. 1.2;
Matches 65; Conservative 59; Mismatches 97; Indels 77; Gaps 16;

QY 13 VKSAIETAGALDLYNKYLDQVTPKTFDETIKELSRFKQYSEASVLVGDIKVLLMDS 72
Db 239 VKSAI--AEQAILAFPKNPEVL-----ETIKIDESYWKNLATTE-----KAFLM-- 280
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASQAQKOILIRILDDGVKKLNEAQK 132
Db 281 -----RTFYQYCN-ENQLHALMDANPELDLSITLEKYLVSRL-----KTINE-NE 325
QY 133 SILTSSQSFNNAAGKLLALDSQLTNDPSEKSSYFQSQVDRIRKEA--YAGAAAGIVAGPF 190
Db 326 NLVKTWETYN---AKDELNDQI---FS-----LENQISRINTDADNFRKSLNSNIED-- 372
QY 191 GLIISYSIAAGVIEGKLIPELNNRLKTVQNPF-----SLSATVKQANK 234
Db 373 --IIEINIAKDLFK-KRIKQKNNSGNLEDLITEENQEIAQDKFLMEDLQQOLEDINK 429
QY 235 DIDAAKLKLATEIAAIGIKTETTTTTFYVDYDDLLSLKGAKKMINTCNEVQQRH 292
Db 430 NLDEIEHQPEDITAKLEELQTK-----YDS-CIRALETTSSELKIQTQVQIFEOAH 477

RESULT 41

US-09-917-254-91
; Sequence 91, Application US/09917254
; Patent No. 6703204
; GENERAL INFORMATION:
; APPLICANT: Mutter, George
; APPLICANT: Baak, Jan
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer
; FILE REFERENCE: B0801/7224(JRV)
; CURRENT APPLICATION NUMBER: US/09/917,254
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/222,093
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 91
; LENGTH: 1857
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-917-254-91

Query Match 6.4%; Score 97.5; DB 4; Length 1857;
Best Local Similarity 21.0%; Pred. No. 4.1;
Matches 66; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

7 EQTEVVVKSIAETADGA-----LDLYNKYLDQVVPKTFDETIKELSRF-----K 51
1058 EQEVTVLKALDETRSHAEQVQEMRQKHAQV-----EELTEQLEQFKRAKANLDKNNK 1111
52 QEYSQSEASVLGDIKULLMDSQKYFEATQTVYEWGCVVTVLLSAYILLDFEYNEKASA 111
1112 QTLEKENADLAGELRVL-----GQAKQEV-----EKKKKLEA 1144
112 QKDLIRILDGVK-----KLNEAQSKLLTSSQFNNASGKLALDLSQLTNDFSEKSS 164
1145 QVQELQSKSDGERARAEKNDKVKHLQNEVESVTGMLNEAEGKAI-----KLAKDVASLSS 1200
165 YFQSQVDRIKREAYAGAAAGIVAGPFGLLIISYIAAGVIE---GKLIPELNNELKTQVNF 221
1201 QLQDTQELLQETROK-----LNVSFKLRQLEERNSLQDQDEMEAKQN- 1246
222 FTSLSATVQKANKDIDAAGKLI---ATEIAAIGBIKTETTTTFYVDYDDLMLSLLKGAA 278
1247 ---LERHISTLNIQLSDSKKKLQDFASTVLEAEGK-----KRFQKEIENLTQOYEKAA 1298
279 --KEMINTCNEYQQ 290
1299 AYDKLEKTKNRLQQ 1312

RESULT 42
US-09-538-092-1084
; Sequence 1084, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuiPatSeqFormat Version 0.9
; SEQ ID NO 1084
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35749
US-09-538-092-1084

Query Match 6.4%; Score 97.5; DB 4; Length 1972;
Best Local Similarity 21.0%; Pred. No. 4.4; Indels 89; Gaps 13;
Matches 66; Conservative 51; Mismatches 108; Indels 89; Gaps 13;
7 EQTEVVVKSIAETADGA-----LDLYNKYLDQVVPKTFDETIKELSRF-----K 51
1173 EQEVTVLKALDETRSHAEQVQEMRQKHAQV-----EELTEQLEQFKRAKANLDKNNK 1226
52 QEYSQSEASVLGDIKULLMDSQKYFEATQTVYEWGCVVTVLLSAYILLDFEYNEKASA 111
1227 QTLEKENADLAGELRVL-----GQAKQEV-----EKKKKLEA 1259
112 QKDLIRILDGVK-----KLNEAQSKLLTSSQFNNASGKLALDLSQLTNDFSEKSS 164
1260 QVQELQSKSDGERARAEKNDKVKHLQNEVESVTGMLNEAEGKAI-----KLAKDVASLSS 1315
165 YFQSQVDRIKREAYAGAAAGIVAGPFGLLIISYIAAGVIE---GKLIPELNNELKTQVNF 221
1316 QLQDTQELLQETROK-----LNVSFKLRQLEERNSLQDQDEMEAKQN- 1361
222 FTSLSATVQKANKDIDAAGKLI---ATEIAAIGBIKTETTTTFYVDYDDLMLSLLKGAA 278

1362 ---LERHISTLNIQLSDSKKKLQDFASTVLEAEGK-----KRFQKEIENLTQOYEKAA 1413
279 --KEMINTCNEYQQ 290
1414 AYDKLEKTKNRLQQ 1427
RESULT 43
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 6.4%; Score 97.5; DB 3; Length 2285;
Best Local Similarity 19.9%; Pred. No. 5.6;
Matches 65; Conservative 58; Mismatches 106; Indels 97; Gaps 16;
9 TVEVVKSAIETADGAL-----DLYNKYLDQVVPKTFDETIKELSRFQOYEQS 57
1091 SVDDIKAAIKEMSDAMQFDSVDVLNGDIFNNTKQVAP---LNDLLEKMAEGKSISANE 1147
58 ASVLVGDIVKVLMDSDQKYFEATQTVYEWGCVV---TQLLSAYILLDFEYNE---K 107
1148 ANTLI-----QKDK--ELAQAISIENGWVVKINRDEVIKQKVKLDAYNDVMVYGNK 1196
108 KASAKDILIRILD-----DGVKKLNEAQSKLLTSSQ-----SFNNASGKLALDLSQ 154
1197 LMKTEVNNAIKTLNADTLRLIDSLKLRKRLDSEALSDLEVKSIINNVAACKEL-KK 1255
155 LTNDFSEKSSYQSOVDRIK-----EAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIP 209
1256 LEEKMLQPGYSNSQIEAMQSVKSALESYISASE-----EATSTQ 1295
210 ELNNRLKTQNFFTSLSATVQKANKDIDAAGKLAETIAAIGETETTTTFYVD-YDD 268
1296 EMNK--QALVEAGTSLNWTDOEK-----ANE-----ETKTSYVVDKYKE 1335
269 LMLSLKGAAG--KEMINTCNEYQORH 292
1336 ALEKVNAEIDKYNKQVNDYPKYSQKY 1361

RESULT 44
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122


```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254--6

Query Match      6.4%; Score 97.5; DB 1; Length 2482;
Best Local Similarity 20.5%; Pred. No. 6.3;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY  7 EQTVVVVKSIAETADGALDLYN----KYLDQVIPW-----KTFDETTELKELSRFKQY 54
Db  1948 QDTLEVLQSSYKNLELELTMDKMSFVEKVNKMTAKETELQREHMAQKTAELQEE 2007
QY  55 SQEASVLVGDIVKVLMD--SODKYFEAT-----QTVVEMGVVTVLLSAY 97
Db  2008 SGEKNRLAGELQLLEIKSSDKQLKELTLENSELKSLDCMHKQDQVEKGGKVEETAEY 2067
QY  98 ILLFDEYNEKKASA-----QKDLIRILDGVKKLNEAQS-----LLTSSQSFN 143
Db  2068 QLRLHE-AEKHQALLDNTNKQVEVEIQTYREKLTSEKCLSSQKLEIDLKSSKEELNN 2126
QY  144 ASGKLALDSQLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYSIAAGVI 203
Db  2127 S-----LKATTTQILEELKTKMDNLKYNQLKENER-----AQGKMKLLI----- 2167
QY  204 EGKLIPELNNRLKTVQNFPTSL-SATVQKANKDIDAACL-KLATEIAAIGE-IKTETETT 260
Db  2168 --KSCQLEBEKEITLQKELSQLQAQEKQKTGTVMDTKVDELTEIKELKETLEETKEA 2225
QY  261 RFYVD-YDDLMLSLK-GAAKGMINT-----CNEYQQRHGK 294
Db  2226 DEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQSKQDSR 2266

RESULT 45
US-09-538-092-1154
; Sequence 1154, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
```

```
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)_(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
; US-09-538-092-1154

Query Match      6.4%; Score 97.5; DB 4; Length 3210;
Best Local Similarity 20.5%; Pred. No. 9.4;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY  7 EQTVVVVKSIAETADGALDLYN----KYLDQVIPW-----KTFDETTELKELSRFKQY 54
Db  2676 QDTLEVLQSSYKNLELELTMDKMSFVEKVNKMTAKETELQREHMAQKTAELQEE 2735
QY  55 SQEASVLVGDIVKVLMD--SODKYFEAT-----QTVVEMGVVTVLLSAY 97
Db  2736 SGEKNRLAGELQLLEIKSSDKQLKELTLENSELKSLDCMHKQDQVEKGGKVEETAEY 2795
QY  98 ILLFDEYNEKKASA-----QKDLIRILDGVKKLNEAQS-----LLTSSQSFN 143
Db  2796 QLRLHE-AEKHQALLDNTNKQVEVEIQTYREKLTSEKCLSSQKLEIDLKSSKEELNN 2854
QY  144 ASGKLALDSQLTNDPSEKSSYFQSDVRIRKEAYAGAAAGIVAGPGLIISYSIAAGVI 203
Db  2855 S-----LKATTTQILEELKTKMDNLKYNQLKENER-----AQGKMKLLI----- 2895
QY  204 EGKLIPELNNRLKTVQNFPTSL-SATVQKANKDIDAACL-KLATEIAAIGE-IKTETETT 260
Db  2896 --KSCQLEBEKEITLQKELSQLQAQEKQKTGTVMDTKVDELTEIKELKETLEETKEA 2953
QY  261 RFYVD-YDDLMLSLK-GAAKGMINT-----CNEYQQRHGK 294
Db  2954 DEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQSKQDSR 2994

RESULT 46
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
```

us-09-993-292b-2.ra1

Wed Jan 5 14:16:58 2005

SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: HUMAN
PCT-US95-16216-1

Query Match 6.4%; Score 97.5; DB 1; Length 3248;
Best Local Similarity 20.5%; Pred. No. 9.6;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY 7 EQTVVVKSAIETADGALDLYN---KYLDQVIPW-----KTFDETTELKLSRPKOEY 54
Db 2676 QDTLEVLQSSYKNLENELELTMDKMSFVEKVNQNTAKETELQREHMEHMAQKTAELQEL 2735

QY 55 SOEASVLGDIKVLMD---SODKYFEAT-----QTVYEWCGVVTQLLSAY 97
Db 2736 SGEKNRLAGEQLLLEIKSSKQQLKELTLENSELKSLDCMHKDOVEKEGKVRERIAY 2795

QY 98 ILLPDEYNEKASA-----QKDIILRIIDDGVKLINEAOKS-----LLTSSQSFNN 143
Db 2796 QLRLHE-AEKHQALLDNTNRYEVEIQTYREKLTSEECUSSOKLEIDLKSSKEELNN 2854

QY 144 ASGKLALDSQLTNDFSEKSSVFQSOVDRIKRAYAGAAAGIVAGPFGLIISYIAAGVI 203
Db 2855 S-----LKATTLQLEBKTKMDNLKYVNLKENER-----AQGKMKLLI----- 2895

QY 204 EGKLIPELNNRLKTVNQFPTSL-SATVQANKDDAACL-KLATEIAAIGE-IKTETETT 260
Db 2896 --KSCQLEEREKEILQKELSQLAAQEKQKTGTVMDTKVDLTTTEIKELKETEKEA 2953

QY 261 RFYVD-YDDLMLSLK-GRAKMWNT-----CNEYQORHGK 294
Db 2954 DEYLDKYCSLLISHEKLEKAKEMLETTQVNLCSQSQSKDSR 2994

RESULT 47
US-09-914-259-66
Sequence 66, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Ratner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 6.4%; Score 97.5; DB 5; Length 3248;
Best Local Similarity 20.5%; Pred. No. 9.6;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY 7 EQTVVVKSAIETADGALDLYN---KYLDQVIPW-----KTFDETTELKLSRPKOEY 54
Db 2676 QDTLEVLQSSYKNLENELELTMDKMSFVEKVNQNTAKETELQREHMEHMAQKTAELQEL 2735

QY 55 SOEASVLGDIKVLMD---SODKYFEAT-----QTVYEWCGVVTQLLSAY 97
Db 2736 SGEKNRLAGEQLLLEIKSSKQQLKELTLENSELKSLDCMHKDOVEKEGKVRERIAY 2795

QY 98 ILLPDEYNEKASA-----QKDIILRIIDDGVKLINEAOKS-----LLTSSQSFNN 143
Db 2796 QLRLHE-AEKHQALLDNTNRYEVEIQTYREKLTSEECUSSOKLEIDLKSSKEELNN 2854

QY 144 ASGKLALDSQLTNDFSEKSSVFQSOVDRIKRAYAGAAAGIVAGPFGLIISYIAAGVI 203
Db 2855 S-----LKATTLQLEBKTKMDNLKYVNLKENER-----AQGKMKLLI----- 2895

QY 204 EGKLIPELNNRLKTVNQFPTSL-SATVQANKDDAACL-KLATEIAAIGE-IKTETETT 260
Db 2896 --KSCQLEEREKEILQKELSQLAAQEKQKTGTVMDTKVDLTTTEIKELKETEKEA 2953

QY 261 RFYVD-YDDLMLSLK-GRAKMWNT-----CNEYQORHGK 294
Db 2954 DEYLDKYCSLLISHEKLEKAKEMLETTQVNLCSQSQSKDSR 2994

RESULT 48
US-09-914-259-66
Sequence 66, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 961
TYPE: PRT
ORGANISM: Bos taurus
US-09-914-259-66

Query Match 6.4%; Score 97; DB 4; Length 961;
Best Local Similarity 22.3%; Pred. No. 1.6;
Matches 67; Conservative 52; Mismatches 125; Indels 56; Gaps 14;

QY 10 VEVVVKSAIETADGALDLYNLYKLDQVIPWKTFTDETELKLSRPKOEYASVLDIKVLL 69
Db 666 LEEELKQKIST-----LKQNEQLQTAV-----TQVVSQIQKHQDQYNL-----LKVQL 708

QY 70 -MDSQDK--YFEATQTVYEWCGVVTQLLSAYILLDFDYNEKKAS-----AKDILIR 118
Db 709 GKDSQHQGPYTDGAQ----MNGVQPEIISR---LREBIEELKSNRELLOLAEKDSLIE 761

Qy 119 ILDDGVKK--LNEAQKSLTSSQFNNASGKLLALDSQLTNDSEKSYFSQSDVRIRKE 176
Db 762 NLKSSQLSPGTNQSATGDSQIAELKQELATLSQL--NSQSVETIKLQTEKQELLQK 820
Qy 177 AYAGAAAGIVAGPGLIISYSIAAGVIEGKL-----IPELNRLKTVQNFFTSLSATVK 230
Db 821 TEAFKAPVPGSSETVIATKTD--VEGRLSALLQETKELKEIKALSERTAIKEQLD 878
Qy 231 OANKIDAALKLATEIAAIGEITETTRFFVYDDMLSLKGAAKMINTCNEYQQ 290
Db 879 SSNSTI-----AILQNEKNKLEVDITDSKEQDQLLV--LLADQDQKIFSLKKNLKE 928

RESULT 49

US-09-489-039A-8352
; Sequence 8352, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8352
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8352

Query Match 6.4%; Score 96.5; DB 4; Length 350;
Best Local Similarity 22.1%; Pred. No. 0.39;
Matches 68; Conservative 46; Mismatches 93; Indels 101; Gaps 17;

Qy 39 TFDTEIKELSRFOEYSQASVL---VGD-IKVLLMDSQDKYFEATQVYEMCGVVTQLL 94
Db 42 TRDVVMAVTLQGYRPNANAQALATQVSDTIGVVMDVSDAFGALVKAVD---TVAQQH 98
Qy 95 SAYILLPDEYNE-KKASQAKDILIR-----ILDD-----GVKKLNEAQ 131
Db 99 QKYVLIGNSYHEAKEKHAIEVLIRQCSALIVHSKALSDDLSDFMHPGMVLINRIV 158
Qy 132 KSLTSSQFNNASGKLLALDSQLTNDSEKSYFSQ-----QVDRIRKEAYAGA--AAGI 185
Db 159 PGVAHRCVGLDNYSGALMA--TRMLLNHGHRIGYLSNHHGIEDDMRECGSKALQEQGI 217
Qy 186 VA-----GP-----FGL--IISY--SIAAGVI-----EGKLIP-- 210
Db 218 IAPDSWIGSGSPDQGGGEAAWVLLGRNLGLTAVFAYNDSWAAGALTTLKDNGIVVPQHL 277
Qy 211 -----LNNRLKTVQNFFTSLSATVKQANKDIDAALKLATEIAAIGEI--KTE 256
Db 278 SLIGFDDIPISRYTDPQLTTVRVPMWSA-----KLATELALLGAAGKLD 322
Qy 257 TETTRFYV 264
Db 323 REATHCFM 330

RESULT 50

US-08-924-629C-6
; Sequence 6, Application US/08924629C
; Patent No. 6403082
; GENERAL INFORMATION:
; APPLICANT: Stiles, Michael E.
; APPLICANT: Vederas, John C.
; APPLICANT: van Belkum, Marius J.
; APPLICANT: Worobo, Randy W.
; APPLICANT: Worobo, Rodney J.

; APPLICANT: Greer, G. Gordon
; APPLICANT: McMullen, Lynn M.
; APPLICANT: Leisner, Jorgen J.
; APPLICANT: Poon, Alison
; APPLICANT: Franz, Charles M.A.P.
; TITLE OF INVENTION: No. 6403082elBacteriocins, Transport and Vector System and Method
; FILE REFERENCE: 660.0005US
; CURRENT APPLICATION NUMBER: US/08/924,629C
; CURRENT FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/026,257
; PRIOR FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 457
; TYPE: PRT
; ORGANISM: lcaad;
US-08-924-629C-6

Query Match 6.4%; Score 96.5; DB 4; Length 457;
Best Local Similarity 21.6%; Pred. No. 0.59;
Matches 73; Conservative 63; Mismatches 113; Indels 89; Gaps 18;

Qy 3 SIFAEOTVEVWKAIB-----TADGALD-----LYNKYLDQVIEPKTFDETIKE 46
Db 39 SLFAKREI--VVRASGEIIPAKVLSDIQSTNSNAIDSNQLTENKVVKGDTLVFTSGNEK 97
Qy 47 LSRFKOEYSQASVLVGDIKVL-----LMSQDKYFEATQVYEMCGVVTQLLSAYILL 100
Db 98 IS--SOLLTQQLNNDRLKSLDTYKQSI VNGRSEFGTDQFGVD-----SLFNGYMAQ 149
Qy 101 FD-----EYNEKK-----ASAQKDLIRILDGKVLNEA---QKSLTSSQS----- 140
Db 150 VDTLTSEFNQSSDKQTADQQAHHQIDVLKQGSKNQQLANYQALTSINSNTKPTNPP 209
Qy 141 ----FNNASGLIALDSQLTNDSEKS-----SYFQSOVDRIR--KEAYAGAAAGIV-AGP 189
Db 210 YQAIYDNYSAQLKS--AQTTDDKQVKQTPALSNVQQIDQLQTTSSYSQIAGITKSGP 267
Qy 190 FGLIISYIAAGVIEGKLI---PELNNRLKTVQNFFTSLSATVKQANKD----- 235
Db 268 LSQSSTLDKIADLKQQQLASAKQEI NDQ-----QQSLDELKAKQSSANEDYQDTVIKAPED 323
Qy 236 -----IDAAKLKLAIEAIAIGEIKTE--TETTRFYVDY 266
Db 324 GILHLATDKTKIKYFPKGTITIAQIYPKLTQKLTALNVEY 361

Search completed: January 5, 2005, 10:59:00
Job time : 20.0526 secs

This Page Blank (uspto)


```
QY 301 PDV 303
Db 303 PDV 305

RESULT 2
C64864
hemolysin E - Escherichia coli (strain K-12)
Name: hemolysin E
Alternate names: hemolysin-inducing protein
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: C64864
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64864
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-305 <BLAT>
A:Cross-references: GB:AB000216; GB:U00096; NID:g1787417; PIDN:AAC74266.1; PID:g1787430
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: hlyE, hpr
C:Function:
A:Description: hemolytic activity
A>Note: pore formation
C:Superfamily: Escherichia coli hemolysin E
C:Keywords: cytotoxic; cytotoxin; hemolysis; transmembrane protein
F:181-197/Domain: transmembrane #status predicted <TMM>
F:123/Active site: Asp #status predicted

Query Match 92.5%; Score 1402; DB 2; Length 305;
Best Local Similarity 91.1%; Pred. No. 1.4e-83;
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTSIFAETVEVWKSATETADGALDLYNKYLDQVVPWKTFTDETIKELSRFKQEYSQASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWQTFTDETIKELSRFKQEYSQASV 62

QY 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKDILIRIL 120
Db 63 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKDILIRIL 122

QY 121 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQTLNDFSEKSSYFQSOVDRIKREAYAG 180
Db 123 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQTLNDFSEKSSYFQSOVDRIKREAYAG 182

QY 181 AAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANKDIDAAK 240
Db 183 AAGVAVGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANKDIDAAK 242

QY 241 LKLTATFAAIGETETETTFYVDYDMLSLKGAACKMINTCNEYQORHGKKTLPFV 300
Db 243 LKLTATFAAIGETETETTFYVDYDMLSLKGAACKMINTCNEYQORHGKKTLPFV 302

QY 301 PDV 303
Db 303 PEV 305

Query Match 91.1%; Score 1387; DB 2; Length 305;
Best Local Similarity 90.1%; Pred. No. 1.3e-82;
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

RESULT 3
E90838
hemolysin E [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90838
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference strain O157:H7
A:Reference number: A99629; MUID:21156231; PMID:11258796
```

```
A:Accession: E90838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <HAY>
A:Cross-references: UNIPROT:Q9REB3; GB:BA000007; PIDN:BA815100.1; PID:g13361141; GSPDB:G1
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECel677
C:Superfamily: Escherichia coli hemolysin E

Query Match 91.6%; Score 1387; DB 2; Length 305;
Best Local Similarity 90.1%; Pred. No. 1.3e-82;
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAETVEVWKSATETADGALDLYNKYLDQVVPWKTFTDETIKELSRFKQEYSQASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWQTFTDETIKELSRFKQEYSQASV 62

QY 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKDILIRIL 120
Db 63 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKDILIRIL 122

QY 121 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQTLNDFSEKSSYFQSOVDRIKREAYAG 180
Db 123 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQTLNDFSEKSSYFQSOVDRIKREAYAG 182

QY 181 AAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANKDIDAAK 240
Db 183 AAGVAVGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANKDIDAAK 242

QY 241 LKLTATFAAIGETETETTFYVDYDMLSLKGAACKMINTCNEYQORHGKKTLPFV 300
Db 243 LKLTATFAAIGETETETTFYVDYDMLSLKGAACKMINTCNEYQORHGKKTLPFV 302

QY 301 PDV 303
Db 303 PEV 305

RESULT 4
E85696
probable pore forming hemolysin hlyE [imported] - Escherichia coli (strain O157:H7, substrain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85696
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <STO>
A:Cross-references: UNIPROT:Q9REB3; GB:AE005174; NID:g12514879; PIDN:AAG56033.1; GSPDB:G1
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: hlyE
C:Superfamily: Escherichia coli hemolysin E

Query Match 91.6%; Score 1387; DB 2; Length 305;
Best Local Similarity 90.1%; Pred. No. 1.3e-82;
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAETVEVWKSATETADGALDLYNKYLDQVVPWKTFTDETIKELSRFKQEYSQASV 60
Db 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWQTFTDETIKELSRFKQEYSQASV 62

QY 61 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKDILIRIL 120
Db 63 LVGDIKVLMDSDQKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKDILIRIL 122

QY 121 DGVVKLNEAOKSLTSSQSFNNASGKLLALDSQTLNDFSEKSSYFQSOVDRIKREAYAG 180
```

Db 123 DDGTTKLNKAEQKSLVSSQSFNAGSKLLALDSQLTNDFSEKSYFOSQVDKIRKERYAG 182
Qy 181 AAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQFFTSLSATVQKANDIDAAK 240
Db 183 AAGVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQFFTSLSATVQKANDIDAAK 242
Qy 241 LKLAETAAIGETITETTFYVDYDMLSLKGAACKMINTCNEYQORHGKKTLEFV 300
Db 243 LKUTTEAAIGETITETTFYVDYDMLSLKGAACKMINTCNEYQORHGKKTLEFV 302
Qy 301 PDV 303
Db 303 PEV 305

RESULT 5
T05634
hypothetical protein F20D10.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T05634
R;Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.R.
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15420
A;Accession: T05634
A;Molecule type: DNA
A;Residues: 1-1496 <BEV>
A;Cross-references: UNIPROT:Q95ZK7; EMBL:AL035538
A;Experimental source: cultivar Columbia; BAC clone F20D10
C;Genetics:
A;Map position: 4
A;Introns: 1042/2; 1212/2; 1232/1; 1263/3; 1349/3
A;Note: F20D10.190

Query Match 7.8%; Score 118; DB 2; Length 1496;
Best Local Similarity 19.6%; Pred. No. 11;
Matches 64; Conservative 67; Mismatches 122; Indels 74; Gaps 13;

Qy 4 IPAEQTVVVKSAIETADGALDLYNK--YLDQVVPKLT-----FDETIKLSRF---KQ 52
Db 153 IEAEKTVKMGKGRDRDVVVQMEEEKSQVEELKWKKEQFKHLEAYEYKLNLFKDSKK 212
Qy 53 EYQSEASVLDGDKVL--LMSQDKYFEATQTVVEMC-GVVTQLLSAYILLFDEYNEKKA 109
Db 213 EWEEEKSLDEIYSLOTKLDVSRISLEDLQKLMCGALTO-----EETRKHIL 263
Qy 110 SAQKDILIRILDGKVLNKAQKSLTSSQSFNAGSKLLALDSQLTNDFSEKSYFQ-- 167
Db 264 EIQVSEFKAYEDAFACQDARTQ-----DLAGKRDEWAEALRQTLMDKDAYFKEM 316
Qy 168 ----SQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEG---KLIPELNNRLKTVQN 220
Db 317 KYENGKLEQENRELGLSLK-----ELQEAITQSGNSALSCLKNKFRLN 362
Qy 221 FFTLSAT-----VKQAKDIDAAKLAETAAIGETITETTFYVDYDMLSLKGAACKMINTCNEYQOR 264
Db 363 IHKNCSANLSKAEAWSQVEEKVVEEINDYKLOLQSKAAALKEVELENCRSSTAKORL 422
Qy 265 DYDDLMLSLKGAACKMINTCNEYQOR 291
Db 423 QYEEISIMFL-----VLSRTVSEAQSR 444

RESULT 6
LECA
hemolysin A - Escherichia coli
C;Species: Escherichia coli
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 09-Jul-2004
C;Accession: A24433; I41280
R;Felmlee, T.; Pellett, S.; Welch, R.A.
J. Bacteriol. 163, 94-105, 1985
A;Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.

A;Reference number: A24433; MUID:85234404; PMID:3891743
A;Accession: A24433
A;Molecule type: DNA
A;Residues: 1-1023 <FEL>
A;Cross-references: UNIPROT:P09983; GB:M10133; GB:M12863; NID:G146377; PIDN:AAA23975.1;
A;Experimental source: strain J96, O4 serotype
R;Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.
Science 266, 1992-1996, 1994
A;Title: Fatty acylation of two internal lysine residues required for the toxic activity
A;Reference number: A5387; MUID:95099325; PMID:7801126
A;Contents: annotation; lysine palmitoylation
A;Note: lysine modification is performed by the hlyC gene product
R;Haertlein, M.; Schiesl, S.; Wagner, W.; Rdest, U.; Kreft, J.; Goebel, W.
J. Cell Biol. 22, 87-97, 1983
A;Title: Transport of hemolysin by Escherichia coli.
A;Reference number: I41280
A;Accession: I41280
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1,'T',3,'V',5,'T',7-44 <RES>
A;Cross-references: GB:M29173; NID:G146337; PIDN:AAA23957.1; PID:G146338
C;Genetics:
A;Gene: hlyA
C;Function:
A;Description: attacks blood cell membranes and causes cell lysis
A;Superfamily: hemolysin A; hemolysin A homology
C;Keywords: calcium binding; cytolysis; exotoxin; hemolysis; lipoprotein; tandem repeat;
F;246-791/Domain: hemolysin A homology <HLYA>
F;723-851/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVITYF]-X)
F;563,689/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match 7.7%; Score 116; DB 1; Length 1023;
Best Local Similarity 20.5%; Pred. No. 9.2;
Matches 69; Conservative 59; Mismatches 111; Indels 98; Gaps 11;

Qy 2 TSIFAEQTVVVK-----SAIETADGALDLYNKYLDQVVPKLTDETIKLSRFKQE 53
Db 155 TALSSWKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207
Qy 54 YSQEASVLD-----VGDIKVLMDSDQKQYFEATQTVVEMCGVVTQLLSAYILL-FDEY 104
Db 208 LNKLGSVLSNTHLNGVGN-KLQNLPLNDINIGAGLDTV---SGILSAISASFILSNADAD 263
Qy 105 NEKKAQAKDILIRILD-----GVKLNKAQKSLTSSQSFNAGSKLLALDSQLTNDFSE 161
Db 264 TGTGAAGVELTTKVLGNVGKISQYIIAQRQAQGLSTS----- 302
Qy 162 KSSYFQSQVDRIRKEAYAGAAAGIVAGPGLIIS---YSIAAGVIEGKLIPELNNRLKT 217
Db 303 -----AAAGLIASVVTLSISPLSLADKFKRANKIEYSQRFKK 344
Qy 218 VQNFPTLSATVQKANDIDAAKLAETAAIGETITETTFYVDYDMLSLKGA 277
Db 345 LGYDGDGSLLAAPHKETGAIDASLTRISLVASVSSGISAATTSV---GAPVSALVGA 400
Qy 278 -----AKWINTCNEYQORHGK 294
Db 401 VTGIISGLEASKQAMFEHVASKQADVIAEWKKGK 437

RESULT 7
T22716
hypothetical protein F55C5.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T22716
R;Harris, B.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19604
A;Accession: T22716
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-622 <WIL>

A;Cross-references: UNIPROT:Q20822; EMBL:Z78198; PIDN:CAB01573.1; GSPDB:GN00023; CESP:FP
A;Experimental source: clone F55C5
C;Genetics:
A;Gene: CESP:F55C5.8
A;Map position: 5
A;Introns: 24/3; 44/2; 147/3; 251/3; 561/3

Query Match 7.6%; Score 115.5; DB 2; Length 622;
Best Local Similarity 21.2%; Pred. No. 5.2;
Matches 68; Conservative 56; Mismatches 109; Indels 87; Gaps 15;

Qy 41 DETIKELSRFKQKQYASVGVGDIKVLMDSQKFEATQTYVCGVVTQLLSAY--- 97
Db 251 DKLISE-RASATSAEVVTVIWGGAQSTVDDE-----KAKQVVOEQTEVELAQCOQTPK 304
Qy 98 --ILLFVEYNEKKASAKDILIRLDGVKLANE-AQKSLTSSQSP-----NNASGKL 148
Db 305 EKMALEF-----EKATADTRDAIDRISDIIRKSSENADTTVLQSIKAYLEFLKMGTAARY 360
Qy 149 LALDSQLTNDFSEKSYFQS-----QVDRIRKEAYAG 180
Db 361 LAI---IDNTKSEKSKPQDLRLLYDSVIEYKEVABIPGADHDKNLQIAFEVKVEYRA 417
Qy 181 AAGAGVAGPGLIISYSTAAGVI-----EGKLIPELNNRLKTVQNFFTSLSATVK 230
Db 418 FRCFYMASSSALHKYSEAAALFRTVSRVQDAEGKL-----KKLKSSSFTITNETQSSLN 472
Qy 231 QANKDIDAAK--LKLAIEAIAIGIKETETETRTFYVDVDDLMLSLGKAQKMNICNEY 288
Db 473 ELRSEVESAKVTVRAARLASAAGDVKTDSELAKI-ID-----KRPLETVNEW 519
Qy 289 QQ---RHQKTLFEVDPVAS 305
Db 520 RQWDRVNSLKDXTIP-VAS 538

RESULT 8
A90394
bPS2 protein homolog (bps2) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: A90394
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
arrett, R.A.; Regan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: A90394
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-587 <KUR>
A;Cross-references: UNIPROT:Q97WH8; GB:AE006641; NID:g13815540; PIDN:AAK42408.1; GSPDB:G
C;Genetics:
A;Gene: bps2

Query Match 7.6%; Score 115; DB 2; Length 587;
Best Local Similarity 20.3%; Pred. No. 5.2;
Matches 61; Conservative 66; Mismatches 103; Indels 70; Gaps 14;

Qy 10 VEVVKSIAETADGALDLYNKYLDQVTPKTFDETI--KELSRFKQYASVGVGDIKV 67
Db 42 ISLTSSIK---AEDLLNVFADSGYVVEALDNKLYRRIKRIRNGLGSEKNLIMDDRA 97
Qy 68 LLMDSQKFEATQTYVCGVVTQLLSA-----YLLFDEYNEKKASAKDILIRILD 121
Db 98 LLL-----TYFSPENR-----LVTQLSGDGNVWFITTSKINEIK--AKKEELQKLT 145
Qy 122 DGVKKLANEAKSLTSSQSFN---ASGKLALDSQLTNDFSEKSYFQSOVDRIKRAY 178
Db 146 AEINARDELQK-----KYNIREIOAKIRAID-----EIDKLEKE-- 181
Qy 179 AGAAAGVAGPGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDA 238

Db 182 RESSNIVAK---TTYTIT--LTRQNKINELINKVKQDELANLEFALKKIEEIQN 234
Qy 239 AKUKLATEIAAIGIKETETRTFYVDYDD---LMLSLGKAQKM-----INTCN 286
Db 235 KESKVSPIKTOLEKEMEEINEKLUKUNDRSELELKVLRVLEVNESDRHLDTCN 294
RESULT 9
T12818
hypothetical protein yonD - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: T12818; F69913
R;Lazarovic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 prophage
A;Reference number: Z17583
A;Accession: T12818
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-478 <LAZ>
A;Cross-references: UNIPROT:O64067; EMBL:AF020713; NID:g3025478; PID:g3025532; PIDN:AAACI
Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 330, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleri
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, Y.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtilis.
A;Title: The complete genome sequence of the Gram-positive bacterium
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69913
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-478 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14031.1; PID:el183560,
A;Experimental source: strain 168
C;Genetics:
A;Gene: yonD

Query Match 7.6%; Score 114.5; DB 2; Length 478;
Best Local Similarity 22.6%; Pred. No. 4.3;
Matches 57; Conservative 41; Mismatches 73; Indels 81; Gaps 12;

Qy 25 DLYNKY-LDOVIPKTFDETIKELSRFKQYASVGVGDIKVLMDSQKFEATQTV 83
Db 268 DVTDTYFVNVYWS--DENSVD-KYFKNYR-----TGDTVSDIDSKTEVF---MT 315
Qy 84 YEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRLDGVKLANEAKSLTSSQSFN 143
Db 316 RNWEEVPEPIQSOL-----NQDEQIK---DLTKQVNNQKDKVIGEQFNT 360
Qy 144 ASGKLALDSQ-----LTNDFSEKSYFQSOVDRIKRAYAGAAAGIVAG 188
Db 361 ASEKLVQNLSEVQGLKYEKHEKHEKLEOKLSEKNEFYKAKFEALNAEE----- 409
Qy 189 PFGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDAUKLATEIA 248
Db 410 -----KFS-----TEEVQNL---IHASVKQ-DEGEKAVLQNTMLV 442
Qy 249 AIGELKTEITET 260
Db 443 DLVSVPTEITNT 454

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19980
A;Accession: T25096
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2823 <W12>
A;Cross-references: EMBL:Z81125; PIDN: CAB03385.1; GSPDB: GN00019; CESP: T22A3.8
A;Experimental source: clone T22A3
C;Genetics:
A;Gene: CESP: T22A3.8
A;Map position: 1
A;Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1988/2; 2760/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 113.5; DB 2; Length 2823;
Best Local Similarity 20.8%; Pred. No. 49;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;
QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESASVLVG-----DIKVLMD 72
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVVGRINRYKEVSEIEKLAVEA 1939

QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104
DB 1940 EDQIAYSRSIEKARSEELMNFEDKEKINNTLAELPDLVEQCNITLL-----YSQLIDEY 1996

QY 105 NEK-----KASAKDILIRILD-----DGVKLNKAEQKS 133
DB 1997 DEEYVQTAGRHAKEVQAK-IVDRFVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055

QY 134 LITSSQSFNAGKLLALDSQLTNDFSEKSYFQSDVRIRKEAYAGAAAGIVAGPGLI 193
DB 2056 AAEASE-----AVSKMLGSGESGDANES--LRQLEKLNKNESSLSN-----2097

QY 194 ISYSIAAGVIE--GKLIPELNNRLKTVQNFPTSLSATVQKANDIDAKKLATEIAAIG 251
DB 2098 VDNSNAVKIVEELKEKDLTRLGHLNELKTSI-----VKRLG 2136

QY 252 EIKTETETFRFYVDYDMLSLKGAKKM-INTCNEYQORHGKKT 296
DB 2137 VIKNEASS---WDDKHDRMHSILKNGAKTAHRSANVKESEGIKT 2179

RESULT 15

T43291
laminin alpha chain - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T43291
R;Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang,
submitted to the EMBL Data Library, June 1998
A;Description: Expression, function and evolution of laminin alpha chains.
A;Reference number: 222397
A;Accession: T43291
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3102 <ZHU>
A;Cross-references: UNIPROT:O45614; EMBL:AF074902; PIDN: AAC26793.1
C;Genetics:
A;Map position: 1
A;Note: lamai/2
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 113.5; DB 2; Length 3102;
Best Local Similarity 20.8%; Pred. No. 56;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;
QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESASVLVG-----DIKVLMD 72
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVVGRINRYKEVSEIEKLAVEA 1939

QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104
DB 1940 EDQIAYSRSIEKARSEELMNFEDKEKINNTLAELPDLVEQCNITLL-----YSQLIDEY 1996

RESULT 13

F87908
protein T22A3.8 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: F87908
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID: 99069613; PMID: 9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: F87908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2823 <STO>
A;Cross-references: UNIPROT:O45614; GB: chr_1; PIDN: CAA15432.1; PID: G3924779; GSPDB: GN00019; CESP: T22A3.8
A;Map position: 1
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 7.5%; Score 113.5; DB 2; Length 2823;
Best Local Similarity 20.8%; Pred. No. 49;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;
QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESASVLVG-----DIKVLMD 72
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVVGRINRYKEVSEIEKLAVEA 1939

QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104
DB 1940 EDQIAYSRSIEKARSEELMNFEDKEKINNTLAELPDLVEQCNITLL-----YSQLIDEY 1996

QY 105 NEK-----KASAKDILIRILD-----DGVKLNKAEQKS 133
DB 1997 DEEYVQTAGRHAKEVQAK-IVDRFVDTRTETENPLKASHAYENIVEALKNATEAVDS 2055

QY 134 LITSSQSFNAGKLLALDSQLTNDFSEKSYFQSDVRIRKEAYAGAAAGIVAGPGLI 193
DB 2056 AAEASE-----AVSKMLGSGESGDANES--LRQLEKLNKNESSLSN-----2097

QY 194 ISYSIAAGVIE--GKLIPELNNRLKTVQNFPTSLSATVQKANDIDAKKLATEIAAIG 251
DB 2098 VDNSNAVKIVEELKEKDLTRLGHLNELKTSI-----VKRLG 2136

QY 252 EIKTETETFRFYVDYDMLSLKGAKKM-INTCNEYQORHGKKT 296
DB 2137 VIKNEASS---WDDKHDRMHSILKNGAKTAHRSANVKESEGIKT 2179

RESULT 14

T23064
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T23064
R;Barlow, K.
submitted to the EMBL Data Library, October 1997
A;Reference number: Z19669
A;Accession: T23064
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2823 <W12>
A;Cross-references: UNIPROT:O45614; EMBL:AL008585; PIDN: CAA15432.1; GSPDB: GN00019; CESP:
A;Experimental source: clone H10E24
R;McMurray, A.

Query Match 7.5%; Score 113.5; DB 2; Length 2823;
Best Local Similarity 20.8%; Pred. No. 49;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;
QY 25 DLYNKYLDOVIPW-KTFDETIKELSRFKQESVQESASVLVG-----DIKVLMD 72
DB 1881 DLKNR-IDVLEQWMDYRETIYDVSKDTADAERMSLVVGRINRYKEVSEIEKLAVEA 1939

QY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104
DB 1940 EDQIAYSRSIEKARSEELMNFEDKEKINNTLAELPDLVEQCNITLL-----YSQLIDEY 1996

Db 1940 EDQIAYSRNIEKARSELNMFEDKEKINWTLAELPDLVEQCQNTLL---YSQLIDEY 1996
QY 105 NEK-----KASAKQDILIRILD-----DGVKKLNEAQAQS 133
Db 1997 DEBYVOTAGHABKLEVQAOK-IVDRVDTRTETENPLKASHAYENIVLVEAKNATEAVDS 2055
QY 134 LRTSSQSFNNASGKLALDSQNTDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPGLI 193
Db 2056 AABASE-----AVSKMLGSESGEDANEEES--JRSQLEKUNESSLSN-----2097
QY 194 ISYISIAAGVIE--GKLIPELNNRLKTVQNFFTSLSATVVKQAKDIDAAKLKLATEIAAIG 251
Db 2098 VDNNSAVKIIVELKKEKKOLTDRLGHLNELKTSI-----VKRLG 2136
QY 252 EIKTETETTRFYDYDDMLSLKGAQKM-INTCNEYQORHGKKT 296
Db 2137 VIKNEASS---WDDKDRMHSILKNGAKTAHRSANVKESEGIKT 2179

RESULT 16
S30834
hypothetical protein YEL043w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 09-Jul-2004
C:Accession: S30834; S50501
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30834
A:Molecule type: DNA
A:Residues: 1-956 <MUL>
A:Cross-references: UNIPROT:P32618; GB:U18779; EMBL:L10830; NID:G603625; PIDN:AAB64999.1
R:Dietrich, F.S.
A:Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.
A:Reference number: S50491
A:Accession: S50501
A:Molecule type: DNA
A:Residues: 1-956 <DIE>
A:Cross-references: EMBL:U18779; NID:G603625; PIDN:AAB64999.1; PID:G603636; MIPS:YEL043w
C:Genetics:
A:Cross-references: SGD:S0000769
A:Map position: 5L

Query Match 7.5%; Score 113; DB 2; Length 956;
Best Local Similarity 24.4%; Pred. No. 13;
Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;

QY 38 KTFDETIELSRPKQEQYSQASVLVGDIVKVLMD-----SQDKYFEATQTVYEWCGVVTQ 92
Db 272 KSLKSNIKSL-----ENSKLLTDLKTEKLKIDKSKKIKRNDMDQKWSQEDTE 322
QY 93 LLS-----AYILLFDBYNEKASAKQDI-----LIRILDDGVKLNKAEAKSLTS---S 138
Db 323 LLSKDTIKYKFKLLNESNASVANINKEIESLQNEISKMEESKRLNASKSLITIVVN 382
QY 139 QSFNN-----ASGKLALDSQNTDFS-EKSSYF-----SQQVDRIKRAYA 179
Db 383 ANVENDKPIASGELSAVLKGL-NDFTLEKNGFLSNAGEEFLSKLNADSSLIKMIKQE--- 438
QY 180 GAAAGIVAGPGLIISYIAA-----GVIEGKLIP-ELNNR-LKT-----V 218
Db 439 -----LSIDQELANWKLQKRLNKKISALENFNEMSLNNRLNLTKLAVQPY 486
QY 219 QNFPTLSAT-----VKQAKNDIDAAKLKLATEIAAIGEI 253
Db 487 KNGKDSLAAATNSNNAEKRRSSGSIQLPLSNMSTRGSI 525

RESULT 17
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: A64465
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, A.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinsteck, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C. A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64465
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1005 <BUL>
A:Cross-references: UNIPROT:Q58718; GB:U67572; GB:L77117; NID:G1591958; PIDN:AAB99331.1
C:Genetics:
A:Map position: REV1273394-1270377
C:Superfamily: hypothetical protein MJ1322

Query Match 7.5%; Score 113; DB 2; Length 1005;
Best Local Similarity 19.7%; Pred. No. 14;
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;

QY 7 EQTVVVKSAIETADGALDLYNKYLDPVWPKTFDE-----TIKELSRF-----KQEQSQ 56
Db 605 DELEDEIKSQLNKFK--NFYQYLSAVSYLSNVDEGIRNRRIKEIENIVSGWNKEKCRE 661
QY 57 EASVL-----VGDIKVLWDSODKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKAS 110
Db 662 ELNKLREDEREINRLKLNELNKEKE-----LIEIENRRSLKFKYKEYLGL 710
QY 111 AQDILIRILDGVKLNKAEAKSLTSSQSFNNASGKLALD-----152
Db 711 TEKLEELKNIKDGLSEI-----YINCNSKILAIIDNIRKYNKEDIYLYNNK 757
QY 153 ----SOLTNDFSKSSYFQSQVDRIKRAYAGAAAGIVAGPGLIISYIAAGVIEGKLI 208
Db 758 ILEVNKEINDIERISYINQKLE-----INYNPE-----EHKKI 792
QY 209 PEL--NNR--LKTQVNFHTSLSATVVKQAKNDIDAAKLKLATEIAAIGEIKTETTRFYV 264
Db 793 KELYENKQELDNVREKQTEIETGIEYLVKQDVESLKARL-----KEMSNLEKKEKLTQFV 848
QY 265 DYDD 268
Db 849 EYLD 852

RESULT 18
S10056
hemolysin A - Escherichia coli plasmid phly152
C:Species: Escherichia coli
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 04-Mar-2000
C:Accession: S10056
R:Hess, J.; Wells, W.; Vogel, M.; Goebel, W. FEMS Microbiol. Lett. 34, 1-11, 1986
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison A:Reference number: S07209
A:Accession: S10056
A:Molecule type: DNA
A:Residues: 1-1024 <HES>
A:Cross-references: EMBL:M14107
C:Genetics:
A:Genome: plasmid phly152
C:Superfamily: hemolysin A; hemolysin A homology
C:Keywords: lipoprotein
F:47-792/Domain: hemolysin A homology <LYA>
F:564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.4%; Score 112.5; DB 2; Length 1024;
Best Local Similarity 20.8%; Pred. No. 15;
Matches 70; Conservative 55; Mismatches 115; Indels 97; Gaps 11;

```

QY 2 TSIFAEQTVVWV-----SAIFADGALDLYNKYLDOVIPWPKTFDETIKELSRFQOE 53
Db 155 TALSSMKIDELIKKQSGGNVSSSELAKASIELINQLVDVTASLN-----NNVNSFSQO 208
QY 54 YSDEASVL-----VGDIKVLMDSDQKCYFEATQVYEWCGVVTQLLSAYILL-FDEY 104
Db 209 LNTLGSVLSNKHNGVGN-KQNLPLNDINAGLDTV---SGILSAISAFILSNADAD 264
QY 105 NEKASAKDILIRIIDD---GVKKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSE 161
Db 265 TRTKAAGVELTTKVLGNVKGISQYIIAQAAGLST-----303
QY 162 KSSYFQSQVDRIRKEAYAGAAAGIVAGPFGLLIS-----YSIAAGVIEGKLIPELNNRLKT 217
Db 304 -----AAAAGLIASAVTLAISPLSIFLSIADKFRANKIEYSQRFK 345
QY 218 VQNFPTSLSATVKQANKDIDAAKILKATEIAAIGEIKETETETTRFYVDYDDMLSLKGA 277
Db 346 LGYDGSLLAFPHKETGALDASLTITSTVLASVSSGISAAATTSV---GAPVSALVGA 401
QY 278 -----AKGMINTCNEVQORHGK 294
Db 402 VTGIISGILEASKQAMFEHVASKMADVIAEWEKKGK 438

RESULT 19
S18199
myosin heavy chain - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 02-Feb-2001
C:Accession: S18199
R:Stewart, A.F.R.; Canoret-Mercado, B.; Perlman, D.; Gupta, M.; Jakovcic, S.; Zak, R.
J. Mol. Evol. 33, 357-366, 1991
A:Title: Structural and phylogenetic analysis of the chicken ventricular myosin heavy chain
A:Reference number: S18199; MUID:92130260; PMID:1774788
A:Accession: S18199
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1039 <SITE>
A:CROSS-references: EMBL:X59552; NID:G62995; PID:CAA42130.1; PID:G62996
A>Note: In the authors' translation 45-Lys is shown after residue 40, and, consequently,
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; nucleotide binding; P-loop
F:402-409/Region: nucleotide-binding motif A (P-loop)

Query Match 7.4%; Score 112; DB 2; Length 1039;
Best Local Similarity 19.9%; Pred. No. 17;
Matches 69; Conservative 67; Mismatches 143; Indels 68; Gaps 11;

QY 1 MTSIFAEQTVVVKSAIETADGALDLYNK--YLDQVIPWKT-FDETIKELSRFQOEYQOE 57
Db 369 MTFMNDLTTQTKLQSENGEFVRQLEEKESLISQLSGKTSFTQOIIEELRRQLEETKS 428
QY 58 ASVLVGDIKVLLMD-----SQDKYFEATQVYEWCGVVTQLLSAYILLPPEYNEK 107
Db 429 KNALAHALQARHDCDLLEQVEEEOEAKELQALSKGNVAQWRTYETDAIQRTPEE 488
QY 108 KASAKDILIRI-----LDDGVKLINEAQKSLTSSQSFNNASGKLLAL 151
Db 489 LEDAKKLLARLQAEAEATEAANKCSLEKAKHRLQNEODMIDLEKANSAAA---SL 545
QY 152 D-----SQJTNDFSEKSYFQSQVDRIRKEAYAGAAAGIVAGPFGLLISYSIAAGVIEG 205
Db 546 DKQGRGFDKIINDWKQYEESSQAELEASQKE-----ARSLSTELFKLKNAYEETLDHLR- 599
QY 206 KLPELNNRLKTQVNFPTSLSATVKQANKDI-----DAAKILKATEIAAIGE 253
Db 600 -----TLKRNKNLQEEISDITNQISEGNKNLHEIKVKQVEQEKSEVQLALEEABGALE 655
QY 254 KTETETTRFYVDYDDMLSLKGAAGKGMINTCNEYQO---RHGKTL 297
Db 656 HEESKTLRF-----QLELSQKADFERKLAEDKDEMQNIRNQORTI 697

```

```

RESULT 20
S21801
myosin heavy chain, neuronal [similarity] - rat
N:Alternate names: myosin II
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: S21801; PNO013; S18134
R:Sun, W.; Chantler, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A:Reference number: S21801; MUID:92235856; PMID:1569576
A:Accession: S21801
A:Molecule type: mRNA
A:Residues: 1-1999 <SUN>
A:CROSS-references: UNIPROT:Q63731; EMBL:X62659
R:Sun, W.; Chantler, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral
A:Reference number: PNO013; MUID:91151356; PMID:1998509
A:Accession: PNO013
A:Molecule type: mRNA
A:Residues: 1914-1998, 'I' <SU2>
A:Experimental source: brain
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:541-575/Region: actin binding #status predicted
F:653-675/Region: actin binding #status predicted
F:836-1999/Domain: coiled coil #status predicted <COI>
F:836-1276/Region: S2
F:1277-1999/Region: light meromyosin
F:125/Modified site: N6,NG,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:693,703/Active site: Cys #status predicted
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 7.4%; Score 111.5; DB 1; Length 1999;
Best Local Similarity 19.6%; Pred. No. 42;
Matches 60; Conservative 49; Mismatches 114; Indels 83; Gaps 11;

QY 7 EQTVVVKSAIE---TADGALDLYNKYLDOVIP--WKTFFDET---IKELSRFQOEYQOE 57
Db 1165 EQEWNILKYLEEAKTHEAQIOEMRQKHSQAEVLAELQETKRVANLEKAKOTLENE 1224
QY 58 ASVLVGDIKVLLMDSDQKCYFEATQVYEWCGVVTQLLSAYILLFDEYNEKASAKDILI 117
Db 1225 RGELANEVKVLQGGRD-----SEHKKKVEAQLOELQ 1257
QY 118 RIIDDGVKLINEAQKSLTSSQSFNNASGKLLALD---SQTNDPSEKSYFQSQVDRIR 174
Db 1258 VKFNEGERRVELADKTKQLQVELDNTVGLSSQDSKSLTKDFSALESQLDQTQELQ 1317
QY 175 KEAYAGAAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTQV---NFFTSLSATVK 230
Db 1318 EEN-----RQKL--SLSTKLQVEDEKKSFRQLEEE 1349
QY 231 QANKDIDAAKILKATEIAAIGEIKETETTRFYVDYDDL-MLSLKGAAGKGMINTCNEYQ 289
Db 1350 EAKINLEK---QIATLHAQVADMKKQME-----DSVCLTAEVKEKQLQKDEGLS 1398
QY 290 QRHGCK 295
Db 1399 QRHEEK 1404

RESULT 21
A90551
conserved hypothetical protein MYPU_3130 [imported] - Mycoplasma pulmonis (strain UAB C)
C:Species: Mycoplasma pulmonis

```

```

398 LLSAVINARQULSIKLLIAESKSEIEEFNIEVAQRVGLGAVAIADUSQNRKISJIVFSDFM 45
271 LSLLLKGAARKRM 282
  |||  |||
458 LALQGNTAPYML 469

RESULT 24
AG1739

```

```

Query Match      7.3%; Score 110; DB 2; Length 821;
Best Local Similarity 20.4%; Pred. No. 17;
Matches 68; Conservative 52; Mismatches 109; Indels 104; Gaps 12;

10 VEVKSAIETADGALDYNKYLDOVTPWKTFTETIKELSRFKQEVSQEASVLVGDIKVL- 68
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 VEKPKRIKELNTEIKVLNG--NQKILOKFPASITVEVNHIGEHENTVTLQONEKILN 430
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 -----LMSODKKFEATQVYEWCGVVTQLLSAYILLFDEYNEKKAQAQDKIL 116
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 DKNVELENKKAELKGNNDKLSEVETYNLDNLSRIVOL-----NDKIEST--DIV 477

117 IRLIDDGVKKLNEAQAQSLTSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSOVDRIKE 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
478 LASKENELDNKLKSLKETLSIKDFNDSD-----LIGQINELISTKNN-LOOQMD----- 526
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

177 AVAGAAAGIVAGPGLIISVSIAGVIEGKLIPELNN---BLKTVQNPFPTSLSATVQKA 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
527 -----DLNNLNDNLLKVVQDCLKIKNEETLKLK 553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

DB 554 EAEIDLSLSEMDLELKQKTSXDDFKWQSKYETVEAKIRNAEVTLENGDIEDLKESK 613

QY 270 -----MLSLKGAAKMINTCNEYQQRHGKKTL 297

DB 614 LHLBETITELNKNVHLENECELEKQKFETSL 646

RESULT 26

MWKWL
myosin heavy chain D [similarity] - Caenorhabditis elegans
N;Alternate names: myosin heavy chain I
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Caenorhabditis elegans
C;Date: 28-Feb-1986 #sequence revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: T21193; S02772; A02993
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19388
A;Accession: T21193
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1938 <WIL>
A;Cross-references: UNIPROT:P02567; EMBL:Z71261; PIDN:CAA95806.1; GSPDB:GN000011
A;Experimental source: clone F21C3
R;Gardner, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19825
A;Accession: T23973
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA

A;Cross-References: none
A;Experimental sources: clone R06C7
R;Dibs, N.J.; Maruyama, I. N.; Krause, M.; Karn, J.
J. Mol. Biol. 205, 603-613, 1989
A;Title: Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy
A;Reference number: S02771; MUID:89178677; PMID:2926820
A;Accession: S02772
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>
A;Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>
A;Cross-References: EMBL:X08065; NID:56785; PIDN:CAA30854.1; PID:56786
R;Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A;Title: Protein structural domains in the *Caenorhabditis elegans* unc-54 myosin
A;Reference number: A93958; MUID:83273600; PMID:6576334
A;Accession: A02993
A;Accession: A02993
A;Molecule type: DNA
A;Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473,
C;Genetics:

Db 924 IAEEREARVEGDKANAKQIEAMKSSVDDSVAAVEEMKKTVAEVERASAEASTNIEALAKT 983

Qy 152 -----DSQLTND-----FSEKSSYFQSQVDRIKREAYAGAAAGIVAGPF 190

Db 984 NIDLALRQDEDHQKQMVNNAKIATTKTQTFADDMASAMASKVEIRAE----- 1029

Qy 191 GLIISYSIAAGVIEGKLIPELNNRLTKVQNFTSLSATVKQANKOIDAA 239

Db 1030 ---IGEDIRASILE-----ETARVEADKTIATHISKLEAQLNDDDISAA 1070

RESULT 28

TI14966

phage lambda-related host specificity protein J - Yersinia pestis plasmid pMT1

C:Species: Yersinia pestis

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: TI14966

R:Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.

A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIMS plasmid

A:Reference number: Z18268; MUID:99043898; PMID:9826348

A:Accession: TI14966

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-1545 <LIN>

A:Cross-references: UNIPROT:Q9ZH03; EMBL:AF074611; NID:G3883003; PID:G3883049; PIDN:AAC8:

C:Genetics:

A:Gene: Y1049

A:Genome: plasmid pMT1

Query Match 7.2%; Score 109.5; DB 2; Length 1545;

Best Local Similarity 19.7%; Pred. No. 41;

Matches 57; Conservative 54; Mismatches 101; Indels 77; Gaps 9;

Qy 2 TSIFAEQTVEVVKSAIEATDAGLDLYNKYLQDVIPWKTDETIKELSRFKQEYSQESAVL 61

Db 849 SSELGQELLEIDA--KASDAQDAINDANKQMEESL--KELDQSVADLDLSKLEDTSGRLQV 904

Qy 62 VGDIKVLIMDSQDKYFEATQTVYEWCGVVTQLL-----SAYIL 99

Db 905 QNDLKNEVSGTLDKVNDALQVEDSNAALVELQETVSEOGKAIGAVEAAHAALDNASAL 964

Qy 100 LFEYNEK-----KASA-QKDILIRILDGVKLNBAQKSLLTSSQSFNNASGKLAL--- 151

Db 965 IAEEREARVEGDKANAKQIEAMKSSVDDSVAAVEEMKKTVAEVERASAEASTNIEALAKT 1024

Qy 152 -----DSQLTND-----FSEKSSYFQSQVDRIKREAYAGAAAGIVAGPF 190

Db 1025 NIDLALRQDEDHQKQMVNNAKIATTKTQTFADDMASAMASKVEIRAE----- 1070

Qy 191 GLIISYSIAAGVIEGKLIPELNNRLTKVQNFTSLSATVKQANKOIDAA 239

Db 1071 ---IGEDIRASILE-----ETARVEADKTIATHISKLEAQLNDDDISAA 1111

RESULT 29

F75008

hypothetical protein PAB1294 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C:Accession: F75008

R:anonymous, Genoscope

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure

A:Reference number: A75001

A:Accession: F75008

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-595 <KAW>

A:Cross-references: UNIPROT:Q9UYA4; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB5050

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1294

QY 262 FYVDYDDLMLSLKGAACKMINTCNEYQORHGK 294
DB 704 SAIEQQTAAIEELRAAQELKDMVGRMRQIVGK 736

RESULT 31

A35254
leukotoxin A - Pasteurella haemolytica (serotype T10)
N:Alternate names: lktA protein
C:Species: Pasteurella haemolytica
C:Date: 10-Aug-1990 #sequence_revision 15-Nov-1996 #text_change 09-Jul-2004
C:Accession: S37145; A35254, S34237; S34235
R:Jainson, A.F.; Aitchison, K.; Donachie, W.
submitted to the EMBL Data Library, September 1993
A:Description: DNA sequence of the leukotoxin A gene from P. haemolytica T10 serotype.
A:Reference number: S37145
A:Accession: S37145

A:Molecule type: DNA
A:Residues: 1-955 <LA1>
A:Cross-references: UNIPROT:P51117; EMBL:Z26247; NID:G400424; PIDN:CAA81206.1; PID:G4004.
R:Highlander, S.K.; Engler, M.J.; Weinstein, G.M.
J. Bacteriol. 172, 2343-2350, 1990
A:Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.
A:Reference number: A35254; MUID:9023688; PMID:2185213
A:Accession: A35254
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 950-955 <HIG>
A:Cross-references: GB:M24197; GB:M34943; GB:M34944
R:Jainson, A.F.; Aitchison, K.D.; Donachie, W.
submitted to the EMBL Data Library, June 1993
A:Description: DNA sequence of the carboxy terminal end of leukotoxin A from the T3 seroi
A:Reference number: S34235
A:Accession: S34237
A:Molecule type: DNA
A:Residues: 745-955 <LA2>
A:Cross-references: EMBL:Z22884; NID:G311828; PIDN:CAA80498.1; PID:G311829
A:Experimental source: serotype T3
A:Accession: S34235
A:Molecule type: DNA
A:Residues: 723-955 <LA3>
A:Cross-references: EMBL:Z22887; NID:G311824; PIDN:CAA80501.1; PID:G311825
A:Experimental source: serotype T10
C:Function:
A:Description: attacks cell membranes and causes cell lysis
C:Superfamily: hemolysin A; hemolysin A homology
C:Keywords: calcium binding; cytotoxicity; exotoxin; hemolysis; lipoprotein; tandem repeat;
F:240-786/Domain: hemolysin A homology <HLIA>
F:718-809/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIVY]-X)
F:718-726/Region: repeat
F:727-735/Region: repeat
F:736-744/Region: repeat
F:745-753/Region: repeat
F:754-762/Region: repeat
F:763-771/Region: repeat
F:772-780/Region: repeat
F:781-789/Region: repeat
F:792-800/Region: repeat
F:801-809/Region: repeat
F:556/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.2%; Score 109; DB 1; Length 955;
Best Local Similarity 18.3%; Pred. No. 24;
Matches 68; Conservative 63; Mismatches 139; Indels 102; Gaps 11;

QY 6 AEQVEVVKSAIETADGALDLYNKVDQVLPWKTFDETIKELSRFKQESQASVLVG-- 63
DB 89 AQTSLGTLQNVILGTERGIVLSAPQLDKLQKNKVGQALGSSSESTAQNFQAKTVLSGVQ 148
QY 64 -----DIKVLMDSDQY-----FEATQTVVWCGVVTQLLSAYILLFDEYNEK 107
DB 149 GNSRTVLGMDLDEALQNESDQLTAKAGLELTNSLIENANSVOTLDAFSEQISQFGSK 208

Query Match 7.2%; Score 109; DB 2; Length 595;
Best Local Similarity 23.0%; Pred. No. 13;
Matches 71; Conservative 55; Mismatches 109; Indels 74; Gaps 18;

QY 24 LDLYNKVDQVLPWKTFDETIKELSRFKQESQASVLVG-DIKVLMDSDQKFEATQT 82
DB 45 LDLASESLDQIEQSNSTSY-ELSRILNKTREDAVIYKRGITKVDLVPFLKLSG 103
QY 83 VYEWCGVVTQLLSAYILLFDEYNEK-----ASAQKDLIRLDDGVKKLNEAQSLLTSS 138
DB 104 I-----GNILKRSFLMYSNFKNEMPELITAMEESLI-IMDDGIK-----EAKLALIRIS 155
QY 139 QSFNNASGKLALD-SQTNDFSEKSYFOSQVDRIRKEAYAGAAIV-----AGPFG 192
DB 156 KIEFSGAGKLDVNDILEKANKLERESSYHKLRL-KSRVTVPRKGLILFVSNANPYVL 214
QY 193 --IISYSTAAG-----VIEGKLPELNRLKTVQNFETSL-----SATVKQANK 234
DB 215 ENVTYFGAKNASSVQILINGSILK-----ANVDNGFFSVNYSFPLPGIYSAVAKSGNL 268
QY 235 D-----IDAALK-----LATBI-AAIGEIKTETETTR-----FVYDDLMLSL 273
DB 269 TSNVTIINVKRKTVFVVPSEVCGRIGETVVISGFLRDLNLGYGVPGRKEIVIDY----- 321

QY 274 LKGAAKQMI 282

DB 322 -KGKATKLI 329

RESULT 30

H75001
methyl-accepting chemotaxis protein (tlpC) PAB1336 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
R:Anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: H75001
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-739 <KAW>
A:Cross-references: UNIPROT:Q3UYF8; GB:A248288; GB:AL096836; NID:G5458960; PIDN:CAB5045
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1336

Query Match 7.2%; Score 109; DB 2; Length 739;
Best Local Similarity 19.2%; Pred. No. 17;
Matches 64; Conservative 71; Mismatches 138; Indels 60; Gaps 11;

QY 1 MTSIFAEQVEVVKSA--TETADGALDLYNKVDQVLPWKTFDETIKELSRFKQESQESA 58
DB 425 VTETRESIGSLVEMANDLEKRANLAQVSKDVTEAI--NQVNEALQVSIQAQOQETI 482
QY 59 SVLVGDIKVLMDSDQKFEATQTVVWCGVVTQLLSAYILLFDEYNEKKAQAQDI--- 115
DB 483 NEITDGHRLVAQTSE-----ESVRAMEBFGAVTEVVS-----IANEGSKGDEALKRIEDI 534
QY 116 --LIRLDDGVKKLNEAQS-----LTSQSQFNASGKLIALDS----- 153
DB 535 QHMSRIETVTSKVAENSRNIEETNVITSIAQTN-----LLALNAAIBAAAGAGEAGRGF 590
QY 154 -----QLTNDFSEKSYFOSQVDRIR---KEAVAGAAAGI-VAGPFGLIISYSIAAG 201
DB 591 AVAQETRKLAEBESQKADNIKSLIKLITKEIKENVEATKEGVSVIGESSETLRDITGYL 650
QY 202 VIEGKLPELNRLKTVQNFETSLSATVQKANKDIDAUKKLATEATAAIGEIKTETTR 261
DB 651 ANIATLLQETSERMTTVE-----QIVRTOBEVDKALREALENLAASAETTASABEVS 703

Qy 108 -----KASAKDILIRILDGKVLNEAKQSLLTSSQSPNNASG 146
Db 209 LQNVKGLGALGDKLKTGGLDKAGLGDVKSRLSLGATAALVLADKDASTAKR-----VG 263
Qy 147 KLLALDSQLTDFSEK--SSYFQSQVDRIKAEVAGAAAGIVA-GPFLGIISYSIAAGV-- 202
Db 264 AGFELANQVGNITKAVSSYILAQ-----RVAAGLSSTGCPVAALIASTAVVAISP 313
Qy 203 IEKLIPELNRLKTVQNF-----TSLSATVQKQANKOIDAAKGLKATEIAAI-- 250
Db 314 LSPAGIADKFDRAKSLNAYRPFKGLGYEGDLSLAIEYQHGTTIDASVTAINALAAIAG 373
Qy 251 -----GEIKTETETTFYVDYDMLSLKGAKKMINTCNE 287
Db 374 GVSAAAAGSVASPIALLVSGITGVISTILOYSK-----QAMFEHVANKIHKIVE 424
Qy 288 YQORHGKKTIFE 299
Db 425 WEXNNGGKNYFE 436
RESULT 32
B70232
hypothetical protein BBG10 - Lyme disease spirochete plasmid G/lp28-2
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: B70232
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kertavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: B70232
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1098 <KLE>
A;Cross-references: UNIPROT:O50733; GB:AB000786; NID:G2690008; PIDN:AAC66075.1; PID:G269
A;Experimental source: strain B31
C;Genetics:
A;Genome: plasmid
Query Match 7.2%; Score 109; DB 2; Length 1098;
Best Local Similarity 20.1%; Pred. No. 28;
Matches 81; Conservative 64; Mismatches 127; Indels 130; Gaps 18;
Qy 2 TSIFAEOT-----VEVKSATETADGALDLYNKY-----LDQV 34
Db 537 SKIFDQUTDENKKILVGKSVNEFNNSYDFVNEYQNLKEKESREREIKTLPHTDQV 596
Qy 35 IPWKTFTETIKELSR-FKQYSQBSASVLVGDIK--VLLMDSQDKYFEATQ----- 81
Db 597 SALQKLNDEINENKNAFVKYKGSFFETLINESNQVVALEKQVNEYKTAldrSFVEAQK 656
Qy 82 -----TVYWCWGVVTVQLLSAYILLFDEYNEKKASAKQDILIRI-----LDD---GVK 125
Db 657 ALQKEITDLEW---ETMLLPA-----KERASAEKQWASKIQAMKFKFVDEHKSQFK 704
Qy 126 KLINEAQK-----SLTSSQSFNNA-----SGKLLALDS-----QL 155
Db 705 KLINEATNTTIQVAEKAQDTTKSLYDSMIDGLNFKNAFMKIDAGKFLNKDGTGESIGEEF 764
Qy 156 TNPFSKSSYFQSQVDRIKAEVAG-----AAAGIVAGPFLGIISYSIAA-----GVI 203
Db 765 HNLINGKDVNKGEGLEKWTQMYESWTKLTAAGAVFGPWGEAVELINGLTDFFWVAIL 824
Qy 204 EGKLIPELNRLKTVQNF-----TSLSATVQKQANKOIDAAKGLKATEIAAI-----GEIKTETE 258
Db 825 KG-----QEKARIKAE-----KKRDELEEKRSSEVELKLEDRFDEIKMRKE 870
Qy 259 TTRFYVDYDMLSLKGAKKMINTCNEYQOR-HGKKTIFE 299

Db 871 KLSLDDDEYTKETIEFLKQAKSQKQISGEIEFQKRLHDVQTEYK 912
RESULT 33
A48669
kinesin-related protein KLP61F - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A48669; B41298
R;Heck, M.M.S.; Persira, A.; Pesavento, P.; Yannoni, Y.; Spradling, A.C.; Goldstein, L.S.B.
J. Cell Biol. 123, 665-679, 1993
A;Title: The kinesin-like protein KLP61F is essential for mitosis in Drosophila.
A;Reference number: A48669; MUID:94043448; PMID:8227131
A;Accession: A48669
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1066 <HEC>
A;Cross-references: UNIPROT:P46863; GB:U01842; NID:G416040; PIDN:AAA03718.1; PID:G416041
R;Stewart, R.J.; Pesavento, P.A.; Woerpel, D.N.; Goldstein, L.S.B.
Proc. Natl. Acad. Sci. U.S.A. 88, 8470-8474, 1991
A;Title: Identification and partial characterization of six members of the kinesin super
A;Reference number: A41298; MUID:92020874; PMID:1924306
A;Accession: B41298
A;Molecule type: DNA
A;Residues: 228-357 <STE>
A;Cross-references: GB:M74428; NID:G157785; PIDN:AAA28655.1; PID:G157786
C;Genetics:
A;Gene: FlyBase:Klp61F
A;Cross-references: FlyBase:FBgn0004378
C;Superfamily: kinesin-related protein KLP61F; kinesin motor domain homology
C;Keywords: ATP; cell division, mitosis; nucleotide binding; P-loop
F;20-362/Domain: kinesin motor domain homology <KMT>
F;103-110/Region: nucleotide-binding motif A (P-loop)
Query Match 7.2%; Score 108.5; DB 1; Length 1066;
Best Local Similarity 21.6%; Pred. No. 30;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;
Qy 24 LDLYNKYLQDVIPWKTFTETIKELSRFKQYSQBSASVLVGDIKVLLMDSQDKYFEATQTV 83
Db 159 LELYNEELCDLL---STDDTTK--IRIFDSTKGSVIIQGLEIPVHSKDDVYKLEKG 213
Qy 84 YEWCGVVTVQLLSAYILLFDEYNEKKASAKQDILIRILDGK-----KLINEAQKSLTSS 138
Db 214 KERKTATTLMNA-----QSSRSHTVFSIVHIRENGIEGEDMLKIGKJMLVDLAGS 265
Qy 139 QSFNNASGK-----LLALDSOLTN--DFSEKSSYFQSQVDRIKAEVAGAA 182
Db 266 ENVSKAGNEKGIKRVETVNIQSLLTLGRVITALVDRAHPVYRESKLTLLQESLGRT 325
Qy 183 -----AGIVAGPFL-----IISYSIAAGVIEGKLIPELNRL--KTVQNFSTLSATVK 230
Db 326 KTSIIATISPCHKDIBETLSTLEYAHRANKIQNK--PEVNQKLTKTTLVKEYTE---EID 380
Qy 231 QANKDIDAAKGLKATEIA--AIGEIKTETETTFYVDYDMLSLK 275
Db 381 KLKRDLMARDKNGIYLAETGYEITLKLESQNLRELNEKMLLLKALK 427
RESULT 34
T44825
hypothetical protein wzc [imported] - Acinetobacter lwoffii
N;Alternate names: protein tyrosine kinase
C;Species: Acinetobacter lwoffii
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44825
R;Nakat, D.; Gutnick, D.L.
submitted to the EMBL Data Library, July 1999
A;Description: Genomic organization of the wzc region of Acinetobacter lwoffii
A;Reference number: Z24856
A;Accession: T44825
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA

QY 272 SLLKGAKKMINTCNEYQORHGK 294
Db 1284 EKLKAKADLEAQLNEVDNHEK 1306

RESULT 37
D84727
Probable RAD50 DNA repair protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: D84727
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umavam, L.; Tallon, L.;
eues, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84727
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1292 <STO>
A;Cross-references: UNIPROT:Q9SL02; GB:AE002093; NID:G4263721; PIDN:AAID5407.1; GSPDB:GN
C;Genetics:
A;Gene: At2g31970
A;Map position: 2
C;Superfamily: RAD50 protein

Query Match 7.1%; Score 107.5; DB 2; Length 1292;
Best Local Similarity 17.2%; Pred. No. 44;
Matches 69; Conservative 77; Mismatches 138; Indels 117; Gaps 13;

QY 2 TSFAQTVVVKSAITADGALDLYNKYLQVWPWT-FDE----- 42
Db 281 STLFKEQORQAALPEENED-----TTEELKEWKSFEERLALLGTRKRMEREMVD 332
QY 43 ---TIKELSRFKQBYSEASVL--VGDIKVLMDSDQKYFEATQTVYEWGVVTTQLLSAY 97
Db 333 TETTISLHNKNTYMLEISKLQTEAEHMLKNERDSTTQNTFFHNLGNVSTFPSTE 392
QY 98 ILLFDEYNKASAKQDILIRLDGQVKLNEAQKSLTSSQSFNNASGKLLALDSQ--- 154
Db 393 VVL--NLTNRIKSRIGLEMDLLDK--KKSNET--ALSTAWDCVMDANDRWKSIEAQKRA 446
QY 155 -----LTDNDFSEKSY-----FQSQ 169
Db 447 KDEIKMGISKRIBEKEITERDSFEFEISTVDVKQTDEREKQVQVELERKTKONSERGFESK 506
QY 170 VDRIRKEAYA-----GAAAGIVAGPFGLIISYSI-----AAGVIEGKLIPELNNR- 214
Db 507 IEQKHIEIYSLEHKIKTLNRRDVMAGDAEDRLTLTRIDECKDRIRGVKGRLPPEKDMKR 566
QY 215 -----LKTQVQFTSLSATVKQANKOITDAAKKLATEIAAIGETETTR-- 267
Db 567 EIVQALRSIEREYDDLSKSRAEKEVNMQLQIEVNNSLFKHNDKTESRKRYIESKLQ 626
QY 268 -----DLMSLLKGAKKMINTCNEYQORHGKTLTFE 299
Db 627 ALKQESVTIDAYPKLESADKRDREYNMANGRMQMF 667

RESULT 38
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR095w; protein YKR415
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C;Accession: S38173; S40647; S31207
R;Baladron, V.; Bailesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38173
A;Molecule type: DNA
A;Residues: 1-1875 <BAL>

A;Cross-references: UNIPROT:Q02455; EMBL:Z28320; NID:G486586; PID:G486587; MIPS:YKR095w
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jin
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromosome
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:G450550; PIDN:CAA51948.1; PID:G450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S31207
A;Molecule type: DNA
A;Residues: 1-300, 'A', 302-1875 <KOE>
A;Cross-references: EMBL:L01992; NID:G171958; PIDN:AAA34783.1; PID:G171959
C;Genetics:
A;Gene: SGD:MLP1
A;Cross-references: SGD:S0001803; MIPS:YKR095w
A;Map position: 11R

Query Match 7.1%; Score 107.5; DB 2; Length 1875;
Best Local Similarity 19.9%; Pred. No. 71;
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;

QY 7 EQTVVVKVKSATETADGALDLYNKYLQVWPWT-FDETI-----KELSRFKQBYSEAS 59
Db 966 EDKISLLKQMFNLNDELQKGMK--EKADFKRISILQNNKEVEAVKSEYSEKLS 1023
QY 60 VLVGDIKVLMDSDQKYFEATQTVYEWGVVTTQLLSAYILLFDEYNKASAKQDI---- 115
Db 1024 KIQND-----LDQOTTYANTQNNYE-----QELQKHADVSKTISELR 1061
QY 116 -LIRILDGQVKLN---BAQKSLTSSQSFNNASGKLLA---LDSQLTNDSEKSSVYQ 167
Db 1062 EQLHTYKQVKTUNLSDQLENALKENKSWSSQKSLLEQLDLSNRIEDLSQNKLLY 1121
QY 168 SQVDRIRKEAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRKTQVQFTSLSA 227
Db 1122 DQI-----QIYTAADKEVNNSTNG-----PGLNNILITLRERDILT 1159
QY 228 TVQANKDIDAAKKLATEIAAIGETETTR-----FYVDYDDL-----LSLL 274
Db 1160 KVTVAERDAKMLRKQISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNL 1219
QY 275 KGAKKMINTCNEYQORHGK 295
Db 1220 R---ESNITLRNELENNNNK 1237

RESULT 39
F72288
methyl-accepting chemotaxis protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: F72288
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: F72288
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-539 <ARN>
A;Cross-references: UNIPROT:Q9X0N0; GB:AE001772; GB:AE000512; NID:G4981693; PIDN:AAD3622
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1146

C;Superfamily: methyl-accepting chemotaxis protein

Query Match 7.0%; Score 106.5; DB 2; Length 539;
Best Local Similarity 19.4%; Pred. No. 17;
Matches 59; Conservative 63; Mismatches 147; Indels 35; Gaps 10;
QY 10 VEVVKSIAETADGAL---DLNKKYLDQVTPWKTFDETIKELSRFKQKYSQEAASVLVGDIK 66
DB 206 IEKIRSKDEIGKAAMAVEKLEIRLLDIITGINKASSEVSSSELSATSEELSANVNSIS 265
QY 67 VLLMDSQDKFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILRILD----- 121
DB 266 EALVSLNKEADENSATLEETASIEELSSR-----ADSNKSAQAAMLESTQVRHEQVEKS 320
QY 122 -DGVKKLNEAQKSLTSSQSFNNASGKLALDSQLT-----NDFSEKSSY--FQSQVDR 172
DB 321 TERIREITERKAHSTREMSSENTKQALNRLLSWAENINSIVDTINSIAEOTNLLALNAIEA 380
QY 173 IRKAYAGAAAGIYAGPFGGLIISYIAA---GVIEGKLIPELNNRLKTQVNFETSL--S 226
DB 381 AR-AGEAGRGFAVVADEIRKLAESKARTQOIGELIKLDEINNSKIVESTASALEET 439
QY 227 ATVQAQKNDIDAAKKLKATEIAAIGETETETETTRFYVDYDDMLSLKGAAGKMMINTCN 286
DB 440 ASLVESIKDV-FESIRIAME-----DVQSRVESVAASTQEQSASLELSAGVTRLTELLN 493
QY 287 EYQQ 290
DB 494 KTRF 497

RESULT 40

AH1369
transmembrane protein [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AH1369
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1369
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-927 <GLA>
A;Cross-references: UNIPROT:O8Y4S2; GB:NC_003210; PIDN:CAD00438.1; PID:gl6411848; GSPDB:
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2360

Query Match 7.0%; Score 106.5; DB 2; Length 927;
Best Local Similarity 20.5%; Pred. No. 33;
Matches 78; Conservative 44; Mismatches 131; Indels 127; Gaps 16;
QY 1 MTSIFAEQTVVVKSAIETADG---ALDLYNKYLDQVTPW---KTFDETIKELSRFKQE 53
DB 160 VTKSYAEAFDKTK---ESGDGFAQAADSGKIKGLVKSGQGNKTIISINLKLADSSLT 216
QY 54 YSQEASVI-----VGDIK-----VLLMDS 72
DB 217 FKDGANTLEVGLKTYTDGVTNTAAAGGDKLNAGVSTLAAAGVPLKDGVAALDGGATKLASG 276
QY 73 QDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILRILDGKVKLNEAQK 132
DB 277 VSTYTSQVDTL---AGGINQAYTGSTALSGLNKNMGS-----VFTLASGITQLNNGQK 327
QY 133 SLTSSQSFNNASGKLIA-----LDSQLTN-----DFSEKSSYFQSQVDRIRKAYAGAA 182
DB 328 SLATGLDLSVDGSKLSAGLKELDGNLTDKQKIAQLKQGMNDLQQGIDQLNQSVNGEDA 387

QY 183 AGIVAGPFGLIISYIAAGVIEGKLIPELNNRL-----KTVQNFETLSAT 228
DB 388 A-----LAKQLAT---LQKSLSDLQGLTFIKSNANFADAEAIKSKINATAGVSAAE 434
QY 229 VKQ-----ANKDIDAOKLATEIAAIGETK-----ETETTRFYVDYDDMLSL 273
DB 435 DKQKIIDAQADLDEKTKGATQVATVEQLQSGLSGLDLAAIQTQVT-----EL 483
QY 274 LKGAAGKMMINTCNBYQQRHG 293
DB 484 QTGVAKISAG-----YQAVHG 499

RESULT 41

C90593
hypothetical protein MYPU 6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: C90593
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis.
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: C90593
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1272 <KUR>
A;Cross-references: UNIPROT:Q98PR8; GB:AL445566; PID:g14090066; PIDN:CAC13824.1; GSPDB:GN
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_6510
A;Genetic code: SGCG

Query Match 7.0%; Score 106.5; DB 2; Length 1272;
Best Local Similarity 23.1%; Pred. No. 50;
Matches 52; Conservative 41; Mismatches 91; Indels 41; Gaps 9;
QY 14 KSAIETADGALDLYNKYLDQVTPWKTFDETIKELSRFKQKYS---QEASVLVGDIKVLLM 70
DB 438 KPRVEPKDETALIFDK-ISKI-----ELKNSQLKQLPSQPKESDLNLSNKLVS 488
QY 71 DSODKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQAQKIDILRILDGKVKLNEA 130
DB 489 DDKNKESELS-----LPQYSISFPKLAGSNNDGTLVDKVIQKQKQEVTKELK 539
QY 131 QKSLTSSQSFNNASGKLALDSQLTNDESEKSSYFQSQV---DRIRKAYAGAAAGIVA 187
DB 540 LTNLLTFEBSLKES-----DFQL--DFSNKKORLASSVWMDKIKESLVVKNKTEN 590
QY 188 GPFGLI-ISYIAAGVIEGKLIPELNNRLKTQVNFETLSATVKQ 231
DB 591 FDFNKYDISYSVSS-----LDEVNGKLKIKMTIFKTKDKRLKE 628

RESULT 42

A35186
salivary agglutinin receptor precursor - Streptococcus sanguis
C;Species: Streptococcus sanguis
C;Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 09-Jul-2004
C;Accession: A35186
R;Demuth, D.R.; Golub, E.E.; Malamud, D.
J. Biol. Chem. 265, 7120-7126, 1990
A;Title: Streptococcal-host interactions. Structural and functional analysis of a Streptococcus agglutinin receptor precursor.
A;Reference number: A35186; MUID:90236997; PMID:2185241
A;Accession: A35186
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1473 <DEM>
A;Cross-references: UNIPROT:P16952; GB:J05418
C;Superfamily: surface antigen spa

Query Match 7.0%; Score 106.5; DB 2; Length 1473;


```

QY      251  GBIKTTETTRFYVD--YDDLMLSLGGAAKMINTCNEYQ-----QRHGKKT 296
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      606  GDFKAYDEASPAVQALLDNAMKDSKRABERKANVDSQYKEIKYQSHAVKT 659

RESULT 47
S54091
hypothetical protein YPR070W - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YP9499.25
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2
C:Accession: S54091; S69058
R:Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54059
A:Accession: S54091
A:Molecule type: DNA
A:Residues: 1-566 <BAD>
A:Cross-references: EMBL:Q12321; EMBL:Z49219; NID:G805025; PID:G8050525
A:Experimental source: strain AB972
R:Couch, J.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of S. cerevisiae cosmid 9513.
A:Reference number: S69057
A:Accession: S69058
A:Molecule type: DNA
A:Residues: 1-566 <COU>
A:Cross-references: EMBL:U51033; NID:G1230676; PID:G1230678; MIPS:YPR0707
C:Genetics:
A:Gene: SGD:MED1
A:Cross-references: SGD:S0006274
A:Map position: 16R

```

QY : : : : :
166 FQNCVDRKRNAGNCRGRI : : : : :
DB 179 FQNCCDFKRTNLNDKFGL : : : : : YILTQG-INGKEVP 211

RESULT 48
T48959
kinesin-like protein - Arabidopsis thaliana
N;Alternate names: protein T15B3.190
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T48959
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unsel, M.; Mewes, H.W.;
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25009
A;Accession: T48959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1229 <OR>
A;Cross-references: UNIPROT:Q9LXV6; EMBL:AL163975; GSPDB:GN00061; ATSP:T15B3.1
A;Experimental source: cultivar Columbia; BAC clone T15B3
C;Genetics:
A;Gene: ATGP:T15B3.190

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:42:38 ; Search time 63.2911 Seconds
(without alignments)
1717.381 Million cell updates/sec

Title: US-09-993-292B-28

Perfect score: 1508

Sequence: 1 MTEIVADKTEVEVKNNAIETA.....TCNEYQKHGKTKLFEVPEV 303

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

A_Geneseq_23Sep04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003s:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1493	99.0	309	2	AAR67645 Haemolysin
2	1402	93.0	305	6	ABG73265 Salmonell
3	1402	93.0	749	6	ABG73266 Salmonell
4	1145	75.9	1035	4	ABG25221 Novel hum
5	126	8.4	1095	6	ABJ25381 Aspergill
6	126	8.4	1277	6	ABJ25981 Aspergill
7	124	8.2	1036	2	ADF07451 Bacterial
8	120.5	8.0	962	2	ADY31646 Human tra
9	120.5	8.0	962	6	ADG52129 E. coli B
10	118	7.8	961	7	ADG52129 E. coli B
11	118	7.8	961	7	ADG52129 E. coli B
12	118	7.8	961	8	ADL99406 Nanosruc
13	118	7.8	1211	5	ADP39975 Staphyloc
14	117.5	7.8	927	5	ABR43563 Hyalurona
15	117.5	7.8	927	5	ABB49720 Listeria
16	116	7.7	885	4	ADG82283 S. epider
17	116	7.7	1129	7	ADF07734 Bacterial
18	115.5	7.7	718	8	ADL05067 M. catarr
19	115	7.6	2056	4	ABM59344 Drosophil
20	114.5	7.6	582	6	ABM68103 Photothab
21	114.5	7.6	1189	6	ABU42996 Protein e
22	114	7.6	515	7	ADC95690 E. faeciu
23	114	7.6	796	5	ABR92030 Herbicida
24	113.5	7.5	1196	7	ADG94317 E. faeciu
25	113	7.5	458	5	ABG80428 Moraxella

26	112.5	7.5	1020	6	ABU30092	Abu30092 Protein e
27	112	7.4	606	2	AAR99673	Aar99673 Receptor
28	112	7.4	631	2	AAR99675	Aar99675 RHAMM 1-2
29	112	7.4	1875	6	ABR53560	AbR53560 Protein s
30	112	7.4	1875	7	ADK64380	Adk64380 Disease t
31	111.5	7.4	808	7	ADF04710	Adf04710 Bacterial
32	111	7.4	435	6	ABR57002	AbR57002 Mouse RHA
33	111	7.4	435	6	ABP97484	Abp97484 Murine RH
34	111	7.4	435	7	ADC02471	Adc02471 Mouse RHA
35	111	7.4	477	6	ABR56999	AbR56999 Mouse RHA
36	111	7.4	477	6	ABP97481	Abp97481 Murine RH
37	111	7.4	477	7	ADC02468	Adc02468 Mouse RHA
38	111	7.4	631	5	ABG50843	Abg50843 Mouse rec
39	111	7.4	794	5	AAU11437	Aau11437 Mouse hya
40	111	7.4	938	6	ABU43682	Abu43682 Protein e
41	111	7.4	1013	4	ABB60789	Abb60789 Drosophil
42	111	7.4	1171	5	ABP73682	Abp73682 Candida a
43	110.5	7.3	304	5	ABP26411	Abp26411 Streptoco
44	110.5	7.3	304	8	ADK99510	Adk99510 Streptoco
45	110	7.3	278	8	ADK99511	Adk99511 Streptoco
46	110	7.3	278	8	ADK99512	Adk99512 Streptoco
47	110	7.3	1163	6	ABU23940	Abu23940 Protein e
48	110	7.3	3187	7	ADE56031	AdE56031 Rat Prote
49	110	7.3	3187	7	ADE56035	AdE56035 Rat Prote
50	109.5	7.3	366	3	AAG23551	Aag23551 Arabidops

ALIGNMENTS

RESULT 1

AAR67645
ID AAR67645 standard; protein; 309 AA.

XX AAR67645;

XX AAR67645;

DT 25-MAR-2003 (revised)

DT 06-JUL-1995 (first entry)

XX Haemolysin regulator.

DE Haemolysin regulator.

XX Haemolysin regulator; tuberculosis; vaccine; therapy.

KW Haemolysin regulator; tuberculosis; vaccine; therapy.

XX Escherichia coli.

OS Escherichia coli.

XX WO9428137-A1.

PN WO9428137-A1.

XX 08-DEC-1994.

PD 08-DEC-1994.

XX 24-MAY-1994; 94WO-US005869.

PF 24-MAY-1994; 94WO-US005869.

XX 24-MAY-1993; 93US-00066830.

PR 24-MAY-1993; 93US-00066830.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA King CH, Shinnick TM, Sathish M;

XX WPI; 1995-022809/03.

XX P-PSDB; AAQ75857.

DR New nucleic acid encoding haemolytic regulator of E. coli - and derived

PT vectors, proteins etc used to transform Mycobacterium cells to produce

PT antituberculosis vaccines providing improved immune response.

XX Claim 9; Page 39-40; 46pp; English.

XX The sequence represents the E. coli haemolysin regulator which may be

CC used as an immunogen in a vaccine composition against tuberculosis

CC (Mycobacterium tuberculosis). (Updated on 25-MAR-2003 to correct PN

CC field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-

CC 2003 to correct PA field.)

XX Sequence 309 AA;

SQ Sequence 309 AA;

```
Query Match      99.0%; Score 1493; DB 2; Length 309;
Best Local Similarity 99.7%; Pred. No. 5.3e-118;
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVVPWQTFDETIKELSRFKQYSQAASV 60
DB 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVVPWQTFDETIKELSRFKQYSQAASV 60

QY 61 LVGDIKTLMDSDQKYFEATQTVVWCGVATQLLAAAYILLPDEYNEKKAQAOKDILIKVL 120
DB 61 LVGDIKTLMDSDQKYFEATQTVVWCGVATQLLAAAYILLPDEYNEKKAQAOKDILIKVL 120

QY 121 DDGITKLNEAKSLVSSQFNASGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180
DB 121 DDGITKLNEAKSLVSSQFNASGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180

QY 121 DDGITKLNEAKSLVSSQFNASGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180
DB 121 DDGITKLNEAKSLVSSQFNASGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180

QY 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240
DB 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240

QY 241 LKLTETIAAIGEIKTETETTRFYDDYDMLSLKKAQKMTNCNEYQKHGKKTLEFV 300
DB 241 LKLTETIAAIGEIKTETETTRFYDDYDMLSLKKAQKMTNCNEYQKHGKKTLEFV 300

QY 301 P 301
DB 301 P 301
```

```
RESULT 2
ABG73265
ID ABG73265 standard; protein; 305 AA.
XX AC ABG73265;
XX DT 17-APR-2003 (first entry)
XX DE Salmonella typhi cytolysin A (ClyA) protein.
XX KW Protein expression vector; expression cassette; export protein;
XX transformed host cell; bacterial cell; immune response; HlyE family;
XX live vector immunogenic composition; animal feed; animal vaccination;
XX food industry; nutritional supplement; biomediation; waste disposal;
XX waste treatment; host bacterium.
XX OS Salmonella typhi.
XX PN US2002146430-A1.
XX PD 10-OCT-2002.
XX PF 23-NOV-2001; 2001US-00993292.
XX PR 22-NOV-2000; 2000US-0252516P.
XX PA (GALE/) GALEN J E.
XX PI Galen JE;
XX DR WPI; 2003-228013/22.
XX DR N-PSDB; ABX15174.
XX PT Novel system for expressing protein of interest, has expression vector
XX with cassette having export protein coding sequence fused to protein
XX coding sequence, host cell transformed with vector, and culturing
XX environment.
XX PS Claim 5; Page 18; 33pp; English.
XX CC The present invention relates to a system for expressing a protein of
XX interest. The system comprises an expression vector having an expression
```

```
CC cassette, where the expression cassette comprises an export protein
CC coding sequence genetically fused to a DNA sequence encoding the protein
CC of interest, a host cell transformed with the expression vector, and a
CC culturing environment for the transformed host cell, where the expression
CC cassette expresses the export protein-protein of interest fusion protein
CC which is exported out of the host cell. The system is useful for
CC expressing a gene in a bacterial cell, by providing an expression vector
CC to a population of untransformed bacterial cells, and expressing the
CC expression cassette, such that the export protein-protein of interest
CC fusion protein is produced and exported into the culture medium. The
CC protein of interest is preferably an antigen and this method is useful
CC for eliciting an immune response from a host against the fusion protein.
CC The export protein may be a member of the HlyE family such as cytolysin A
CC (ClyA). The system is useful for the production of recombinant proteins
CC inside a bacterial host cell, in a bioreactor, and various devices that
CC facilitate the growth of bacteria. The system is also useful for
CC providing an animal antigenic material against which an immune response
CC may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can
CC be used as a live vector immunogenic composition capable of facilitating
CC the generation of an immune response in an animal. The system is also
CC useful for preparing immunogenic compositions for vaccinating animals,
CC and for use in the food, nutritional supplement, animal feed,
CC biomediation, waste disposal, and waste treatment industries. The system
CC is very advantageous for protein expression, as proteins that are toxic
CC to host bacterium can also be expressed. A population of recombinant host
CC cells can be transformed with the expression vector, and then the
CC population of bacterial host cells can be maintained in culture and used
CC to produce proteins without having to harvest and lyse the host cells.
CC The present represents Salmonella typhi ClyA protein
XX Sequence 305 AA;
```

```
Query Match      93.0%; Score 1402; DB 6; Length 305;
Best Local Similarity 91.1%; Pred. No. 2.7e-110;
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVVPWQTFDETIKELSRFKQYSQAASV 60
DB 1 MTSIFAEQTVVVKNAIETADGALDLYNKYLDQVVPWQTFDETIKELSRFKQYSQAASV 60

QY 61 LVGDIKTLMDSDQKYFEATQTVVWCGVATQLLAAAYILLPDEYNEKKAQAOKDILIKVL 120
DB 61 LVGDIKTLMDSDQKYFEATQTVVWCGVATQLLAAAYILLPDEYNEKKAQAOKDILIKVL 120

QY 121 DDGITKLNEAKSLVSSQFNASGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180
DB 121 DDGITKLNEAKSLVSSQFNASGKLALDLSQTLNDFSEKSYFOSQVDKIRKEAYAG 180

QY 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240
DB 181 AAAGVVAGPFGLLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQANKDIDAAK 240

QY 241 LKLTETIAAIGEIKTETETTRFYDDYDMLSLKKAQKMTNCNEYQKHGKKTLEFV 300
DB 241 LKLTETIAAIGEIKTETETTRFYDDYDMLSLKKAQKMTNCNEYQKHGKKTLEFV 300

QY 301 PEV 303
DB 301 PDV 303

RESULT 3
ABG73266
ID ABG73266 standard; protein; 749 AA.
XX AC ABG73266;
XX DT 17-APR-2003 (first entry)
XX DE Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.
XX KW Protein expression vector; expression cassette; export protein;
XX transformed host cell; bacterial cell; immune response; HlyE family;
```

cytolysin A; ClyA; recombinant protein; bioreactor; bacterial growth; live vector immunogenic composition; animal feed; animal vaccination; food industry; nutritional supplement; biomediation; waste disposal; waste treatment; host bacterium; sacB; mutant; muten; lethal levansucrase.

Salmonella typhi.
Bacillus subtilis.
Synthetic.
Chimeric.

US2002146430-A1.
10-OCT-2002.
23-NOV-2001; 2001US-00993292.
22-NOV-2000; 2000US-0252516P.
(GALE/) GALEN J E.
Galen JE;
WPI; 2003-228013/22.
N-PSDB; ABX15191.

Novel system for expressing protein of interest, has expression vector with cassette having export protein coding sequence fused to protein coding sequence, host cell transformed with vector, and culturing environment.

Example 3; Page 25-28; 33pp; English.

The present invention relates to a system for expressing a protein of interest. The system comprises an expression vector having an expression cassette. Where the expression cassette comprises an export protein coding sequence genetically fused to a DNA sequence encoding the protein of interest, a host cell transformed with the expression vector, and a culturing environment for the transformed host cell, where the expression cassette expresses the export protein-protein of interest fusion protein which is exported out of the host cell. The system is useful for expressing a gene in a bacterial cell, by providing an expression vector to a population of untransformed bacterial cells, and expressing the expression cassette, such that the export protein-protein of interest fusion protein is produced and exported into the culture medium. The protein of interest is preferably an antigen and this method is useful for eliciting an immune response from a host against the fusion protein. The export protein may be a member of the HlyE family such as cytolysin A (ClyA). The system is useful for the production of recombinant proteins inside a bacterial host cell, in a bioreactor, and various devices that facilitate the growth of bacteria. The system is also useful for providing an animal antigenic material against which an immune response may be mounted. The obtained recombinant bacterium (e.g. Salmonella) can be used as a live vector immunogenic composition capable of facilitating the generation of an immune response in an animal. The system is also useful for preparing immunogenic compositions for vaccinating animals, and for use in the food, nutritional supplement, animal feed, biomediation, waste disposal, and waste treatment industries. The system is very advantageous for protein expression, as proteins that are toxic to host bacterium can also be expressed. A population of recombinant host cells can be transformed with the expression vector, and then the population of bacterial host cells can be maintained in culture and used to produce proteins without having to harvest and lyse the host cells. The present sequence represents Salmonella typhi ClyA-Bacillus subtilis lethal levansucrase (encoded by sacB gene) fusion protein. The fusion protein is used to investigate the verastility of ClyA as a fusion partner to export heterologous antigens out of S. typhi

Sequence 749 AA;

Query Match 93.0%; Score 1402; DB 6; Length 749;
Best Local Similarity 91.1%; Pred. No. 9.1e-110;
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTRIVADKTVEVVKNAIETADGALDLNKNYLDQVIMQTFDETIKELSRFKQYSQASV 60
DB 1 MTSIFAEQTVVVKSAIETADGALDLNKNYLDQVIMQTFDETIKELSRFKQYSQASV 60
QY 61 LVGDIKTLLMDSQDKYFEATQTVYVCGVATQLLAAVILLPDEYNEKASAKDKILIKVL 120
DB 61 LVGDIKVLMLDSQDKYFEATQTVYVCGVATQLLAAVILLPDEYNEKASAKDKILIKVL 120
QY 121 DGGITKLNEAQSLLVSSQSFNNASGKLLALDSQTNDFSEKSYFQSQVDKIRKEAYAG 180
DB 121 DGGVKKLNEAQSLLVSSQSFNNASGKLLALDSQTNDFSEKSYFQSQVDKIRKEAYAG 180
QY 181 AAAGVAGPFLIISYSIAAGVVEGKLIPELKNKLSVQNFFTLSLTVQANKDIDAAK 240
DB 181 AAAGVAGPFLIISYSIAAGVVEGKLIPELKNKLSVQNFFTLSLTVQANKDIDAAK 240
QY 241 LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLKEAAKMKMINTCNEYQQRHGKKTLEFV 300
DB 241 LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLKEAAKMKMINTCNEYQQRHGKKTLEFV 300
QY 301 PEV 303
DB 301 PDV 303

RESULT 4
ABG25221
ID ABG25221 standard; protein; 1035 AA.
XX AC ABG25221;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #25212.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS89408.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 55580; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1035 AA;

Query Match 75.9%; Score 1145; DB 4; Length 1035;
Best Local Similarity 100.0%; Pred. No. 8.4e-88;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 MDSQKYFEATQVYEWGCVATQLLAAVILLFDEYNEKKAQAQKIDILIKVLDGDTKLE 129
DB 1 MDSQKYFEATQVYEWGCVATQLLAAVILLFDEYNEKKAQAQKIDILIKVLDGDTKLE 60

QY 130 AQSLLVSSQSFNNASGKLLALDSQLTNDSEKSYFOSQVDKIRKEAYAGAAAGVWAGP 189
DB 61 AQSLLVSSQSFNNASGKLLALDSQLTNDSEKSYFOSQVDKIRKEAYAGAAAGVWAGP 120

QY 190 FGLIISYSIAAGVWEGKLIPELKNKLSQVQFFTLTSLNTVKQAKDIDAAKLKLTTEIAA 249
DB 121 FGLIISYSIAAGVWEGKLIPELKNKLSQVQFFTLTSLNTVKQAKDIDAAKLKLTTEIAA 180

QY 250 IGEIKTETTRFYDDYDMLSLKEAKKQKNTCNEYQKHGKTLFE 299
DB 181 IGEIKTETTRFYDDYDMLSLKEAKKQKNTCNEYQKHGKTLFE 230

RESULT 5
ABJ25381
ID ABJ25381 standard; protein; 1095 AA.
XX
AC ABJ25381;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #39.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
FN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroehkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX

PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or inhibit formation of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterization, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 1095 AA;

Query Match 8.4%; Score 126; DB 6; Length 1095;
Best Local Similarity 21.4%; Pred. No. 0.2;
Matches 73; Conservative 49; Mismatches 73; Indels 146; Gaps 15;

QY 4 IVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFQKQYSAASV--L 61
DB 281 IKKEKIDIEATNALVPVDEKVDITRKVER-----FASRIAEG--KERDQAANVKOL 332

QY 62 VGDIKTLMDSQKYFEATQVYEWGCVATQLLAAVILLFDEYNEKKAQAQKIDILIKVLD 121
DB 333 EKDLK--VVEKAQAQWEA-----EW----- 350

QY 122 DGITKLEAQSLLVSSQSFNNASGKLLALDSQ---LTNDFSEKSYFOSQVDKIRKEA 177
DB 351 -----HKTMSKGGQLSESDQOEYRMLKEEYKRSASQAQINLDNLKROR 394

QY 178 YAGAAAGVWAGPFLIISYSIAAGVWEGKLIPELKNK-----LKSQVQFFTT----- 224
DB 395 KTEAEA-----YN-----SLKSKFDSQTEWOLKSVENDTQTLTERKS 430

QY 225 -LSNTVKQAKDIDAAKLK---LTTEIAAIGEIKTETTRFYDDYDMLSLKEA--- 277
DB 431 ALNDTVKTTSKEDIRKKKELNALTSELRISQWTELE-----EKQVVLKLEADDG 484

QY 278 -----AKKMIT-----CNEYQKHOK 294
DB 485 KQTERELRAKELISTLKRIPPGVKGRVSDLCRPQKQKYAE 525

RESULT 6
ABJ25981
ID ABJ25981 standard; protein; 1277 AA.
XX
AC ABJ25981;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #639.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX

OS Aspergillus fumigatus.
 PN WO200286090-A2.
 XX 31-OCT-2002.
 XX 23-APR-2002; 2002WO-US013142.
 XX 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX (ELIT-) ELITRA PHARM INC.
 XX JIANG B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT fumigatus, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page; 175pp; English.
 XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This sequence represents a protein of one of the essential genes
 CC of *Aspergillus fumigatus* of the invention
 XX SQ Sequence 1277 AA;
 Query Match 8.4%; Score 126; DB 6; Length 1277;
 Best Local Similarity 21.4%; Pred. No. 0.24;
 Matches 73; Conservative 49; Mismatches 73; Indels 146; Gaps 15;
 4 IVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIELSRFKQESQAASV--L 61
 323 IKKEDLEAEATNALVPVDEKVDITRKVER-----FASRTAETG--KERDQAANYKQL 374
 62 VGDITKLMDSDQKYFATQTVYEWCGVATOLLAAAYILLFDEYNEKKAQAQKILIKVLD 121
 375 EKDLK--VVEKAQAQWEA-----EW----- 392
 122 DGITKLNEAQSLLVSGSOFNAGSKLLALDSQ-----LTNDFSKSSYFQSQVDKIRKEA 177
 393 -----HKTMSNKGQLSSESQDQYKMLKEVRKSRSSAEQINLDNLKQR 436
 178 YAGAAAGVAGPFGLLISYSYTAAGVGEKLTPELKNK-----LKSQNEFFT----- 224

437 KTEAEA-----YN-----SLKSKFDSTEWQLKSVENDTQTLTERKS 472
 225 -LSNTVKQANKOIDAANKL---LTTEIAAIGEIKTETETFRFYVDYDMLSLKEA--- 277
 473 ALNDTVKTSKEIDRKKKELNALTSELRISQMRTELE-----EKVQVVLKLLLEADDG 526
 278 -----AKKMIT-----CNEYOKRHGK 294
 527 KKQTERELRAKELISTLKRIFPGVGRVSDLCRPKQKKYAE 567
 RESULT 7
 ADF07451
 ID ADF07451 standard; protein; 1036 AA.
 AC ADF07451;
 XX 12-FEB-2004 (first entry)
 DT Bacterial polypeptide #3564.
 DE *Proteus mirabilis* infection; bacterial infection; antibacterial;
 XX immunostimulant.
 KW *Proteus mirabilis*.
 OS *Proteus mirabilis*.
 XX US6605709-B1.
 PN 12-AUG-2003.
 XX 05-APR-2000; 2000US-00543681.
 PF 09-APR-1999; 99US-0128706P.
 PR (GENO-) GENOME THERAPEUTICS CORP.
 XX Breton GL;
 XX WPI; 2003-895291/82.
 DR N-PSDB; ADF03279.
 XX New *Proteus mirabilis* polypeptides and polynucleotides, useful as
 PT reagents for diagnosis of bacterial disease, as components of
 PT antibacterial vaccines, as targets for antibacterial drugs, or as
 PT biocontrol agents for plants.
 XX Disclosure; SEQ ID NO 7736; 870pp; English.
 XX The invention relates to new *Proteus mirabilis* polypeptides and
 CC polynucleotides. The invention also relates to antibodies against the
 CC polypeptides, methods for producing the polypeptides, a method of
 CC generating vaccines for immunising an individual against *P. mirabilis*, a
 CC method for evaluating a compound for the ability to bind a *P. mirabilis*
 CC polypeptide and a method for screening test compounds for anti-bacterial
 CC activity. The polypeptides and polynucleotides are useful as molecular
 CC targets for diagnosing, preventing and treating pathological conditions
 CC resulting from bacterial infection, as reagents for diagnosis of
 CC bacterial diseases, as components of antibacterial vaccines, as targets
 CC for antibacterial drugs or as bio-control agents for plants. This
 CC sequence represents a *Proteus mirabilis* polypeptide of the invention.
 XX SQ Sequence 1036 AA;
 Query Match 8.2%; Score 124; DB 7; Length 1036;
 Best Local Similarity 21.5%; Pred. No. 0.27;
 Matches 82; Conservative 56; Mismatches 127; Indels 116; Gaps 16;
 4 IVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIELSRFKQESQAASV--L 61
 178 IKKQKNGENVSSS-ELAKASIDLINQVLD-----TASSLNNNNISAFSQNLKGLSVLSN 230
 62 -----VGDITKLMDSDQKYFATQTVYEWCGVATOLLAAAYILL-FDEYNEKKAQAQK 114

Db 231 TKHLNGVN-KLONPLNLDKLTGLDVT-...SGILSAISAFILSNADADTGKAAAGVE 286
QY 115 ILIKVLD-...GITKLEAQKSLVSSQSFNNASGKLLALDSQL--T 156
Db 287 LTTKVLGVNGKAVSQYILAQVAGLS- TSAASAGLIASAVTLAISPLSLATADQFKRA 345
QY 157 NDFSEKSYFQ-...SOVDKIRKEA-...YAGAAAGVVAGPFG 192
Db 346 NKIEYSQRFKFGVGDGSLAFAFRKETGAIDASLTINTALGTISAGISAASLASLIGA 405
QY 193 IISVSIAR--GVVEGLIPELKNKLSQVQFFTLSTNV-... 229
Db 406 PISALVGAITGIISGLEASQSMFEHVAN---RWANTIAEWKTHGNFENGYDARHS 462
QY 230 -----KQANKDIDAALKLTT-----EIAAIGEIKTETETTRFFVYDDLM 270
Db 463 AFLEDNFKLLSQYNKEYSVERSVLITQHWDELIGELASVTNGAKTILSGKSYIDY---- 518
QY 271 LSLKKAARKMINTCNEYQKR 291
Db 519 ----YEEGRLEKKEPQK 535

RESULT 8

AAY31646
ID AAY31646 standard; protein; 962 AA.

AC AAY31646;

DT 02-NOV-1999 (first entry)

XX Human transport-associated protein-8 (TRANP-8).

XX Transport-associated protein; TRANP; nuclear pore; nuclear transport;
KW vesicle trafficking; cancer; cystic fibrosis; multidrug resistance;
KW hypercholesterolaemia; diagnosis; treatment.

OS Homo sapiens.

Location/Qualifiers

Key 18
Modified-site /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 34 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 74 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 81 /note= "O-phosphorylated by tyrosine kinase"
FT FT
Modified-site 91 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 101 /note= "N-glycosylated"
FT FT
Modified-site 123 /note= "N-glycosylated"
FT FT
Modified-site 129 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 243 /note= "N-glycosylated"
FT FT
Modified-site 336 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 410 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 451 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 453 /note= "N-glycosylated"
FT FT
Modified-site 585 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 631 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 632 /note= "O-phosphorylated by casein kinase II"
FT FT
Modified-site 717 /note= "O-phosphorylated by casein kinase II"
FT FT

FT Modified-site /note= "O-phosphorylated by casein kinase II"
FT 754 /note= "O-phosphorylated by casein kinase II"
FT Modified-site /note= "O-phosphorylated by casein kinase II"
FT 758 /note= "O-phosphorylated by casein kinase II"
FT Modified-site /note= "O-phosphorylated by casein kinase II"
FT 780 /note= "O-phosphorylated by casein kinase II"
FT Modified-site /note= "O-phosphorylated by casein kinase II"
FT 844 /note= "O-phosphorylated by casein kinase II"
FT Modified-site /note= "N-glycosylated"
FT 882 /note= "O-phosphorylated by casein kinase II"
FT Modified-site /note= "O-phosphorylated by casein kinase II"
FT 890 /note= "O-phosphorylated by casein kinase II"
FT Modified-site /note= "O-phosphorylated by casein kinase II"
FT 902

W09941373-A2.

19-AUG-1999.

05-FEB-1999; 99WO-US002527.

11-FEB-1998; 98US-00021764.

(INCY-) INCYTE PHARM INC.

Au-Young J, Hillman JL, Lal P, Guegler KJ, Corley NC, Yue H;
Bandman O, Baughn MR;

WPI; 1999-508646/42.

N-PSDB; AAZ11738.

Human TRANP coding sequences, used to treat transport disorders and cancer.

Claim 1; Page 74-77; 87pp; English.

This sequence represents human transport-associated protein-8 (TRANP-8). The DNA sequence was first identified in a human colon tissue cDNA library. The full-length cDNA was derived from a series of overlapping and/or extended cDNA sequences and is a consensus. TRANP-1 to 9 (AAY31639 -Y31647) are a novel group of proteins with chemical and structural homology that are involved in the transport of molecules, either intracellularly or to the extracellular environment. Examples of such disorders include cystic fibrosis, multidrug resistance, hypercholesterolaemia and certain forms of diabetes mellitus. Defective nuclear transport may play a role in cancer. For example, the BRCA1 protein, associated with familial breast cancer, is normally located in the nucleus via nuclear pore complexes, but is aberrantly located in the cytoplasm in breast cancer cells. In other cancers, cells can secrete excessive amounts of hormones e.g. cancers of the adrenal medulla can secrete excessive amounts of adrenaline and noradrenaline, leading to hypertension. TRANP is expressed in cancer cells, and transport disorders result from either excessive or insufficient molecular transport. Anti-TRANP antibodies and nucleic acids encoding TRANP can be used as diagnostic tools for such disorders. TRANP antagonists can be used to treat or prevent a cancer associated with increased TRANP expression. Anti-TRANP antibodies can be used directly as an antagonist or as a targeting mechanism for drugs. Alternatively, a TRANP antisense nucleotide can be used to treat cancers. A TRANP agonist or expression vector may be used to treat a disorder caused by reduced transport of biologically active molecules

Sequence 962 AA;

Query Match 8.0%; Score 120.5; DB 2; Length 962;

Best Local Similarity 23.7%; Pred No. 0.48;

Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTVEVVKNAIETADGALDLY-NKYLDQVIVPWTFDETIKLSRFKQYSQAASVLVCDI 65

635 DKKEEVKKTLEQHDNIVTHYKNMIREQDLQLELRQVSTLKQCNQEQLTAVTQVQSQI 694

Db

15-APR-2003; 2003US-00414532.
16-APR-2002; 2002US-0372710P.
(CURT/) CURTISS R.
(KANG/) KANG H Y.
Curtiss R, Kang HY;
WPI; 2004-399655/37.
N-PSDB; ADO52074.
New vaccine comprising a live attenuated strain of pathogenic gram-
negative bacteria, useful in eliciting a Th2-type immune response in a
vertebrate against pathogens, e.g., helminths, fungi, viruses, protozoans
or bacteria.
Example 5; Fig 13; 94pp; English.
The invention relates to immunogenic compositions and vaccines comprising
a live attenuated strain of pathogenic gram negative bacteria that
secretes an antigen. The vaccine is useful in eliciting a Th2-type immune
response in a vertebrate against pathogens, e.g., helminths, fungi,
viruses, protozoans or bacteria. The present sequence is Escherichia coli
beta-lactamase signal peptide (bla SS)-Streptococcus pneumoniae
CC pneumococcal surface protein A (PspA)EF5668-pspARx1 chimeric protein.
CC This sequence is used in the exemplification of the invention. Note: This
CC sequence is stated to be similar to the sequence shown in the sequence
CC listing, however these sequences differ.
XX
SQ Sequence 627 AA;
Query Match 7.8%; Score 118; DB 8; Length 627;
Best Local Similarity 22.5%; Pred. No. 0.44; Mismatches 128; Indels 94; Gaps 15;
Matches 79; Conservative 50;
QY 11 EVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKOEYSQAASVLVGDITLLM 70
Db 173 DYKEGLEQA-----ITNKAELATTQONIDKTQDLEAELEKVLATLDPGKT--Q 225
QY 71 DSQDKVFEATQVYVCGVATQLLAAYILLPDEYNEKASAKDILIKLVLDGDTKL--- 127
Db 226 DELDK--EAEA-----ELNEKVEALQNO--VAELEELSKLEDN 261
QY 128 -----NEAKSLVSSQSFNNASGKL-----LALD-----SOLTNDPSEKSSY 165
Db 262 LKDAETLQSPVASQSKAEKDYDAKKDAKNKAKKAVEQAQALDDAKAAQKYDEQKTE 321
QY 166 FQSQVDKIRKEAVAGAAAGVAGPFGILISYSTAAGVVEGKLIPELKNKLKSVQNFPTT- 224
Db 322 EKALEKAASEEMDKAAVAQVQAYLAYQQAATDKAAKDAADKMIDEAKGEEEAKTFTV 381
QY 225 -----LSNTVQA-----NKDIDAAKLKLTETIAAIGEIKTE-----TE 258
Db 382 RAMVVPPEQLAETFKKSEAKQAPELTKLSEAKKL--EAAALEKAASEEMDKAAVA 440
QY 259 TTRFVVDYDILMLLSLLEAKAKMINTCNEYQKRGK- KTLFE-----VPE 302
Db 441 VQAYLAYQQAATDKAAKDAADKMI---DEAKGEEEAKTFTVTRAMVPE 488
RESULT 11
ADB67135
ID ADB67135 standard; protein; 961 AA.
XX
AC ADB67135;
XX
DT 04-DEC-2003 (first entry)
DE General vesicular transport factor p115 SEQ ID NO:155.
XX staged assembly; nanostructure; peptide nucleic acid; PNA;
KW structural reinforcement; aerogel; paper; plastic; cement;
KW
tensile strength; identification marker; anti-counterfeiting marker;
enzyme support; catalyst support; assembly scaffold; nanowire;
nanocircuit; molecular sieve; molecular filter; biosensor.
Bos taurus.
WO2003072829-A1.
04-SEP-2003.
21-FEB-2003; 2003WO-US005390.
21-FEB-2002; 2002US-00080608.
(NANO-) NANOFAMES INC.
Hyman PL, Goldberg EB;
WPI; 2003-721788/68.
Staged assembly of nanostructures, useful e.g. in biosensors or as
catalyst supports, using assembly units derived from peptide nucleic
acids.
Disclosure; Page 58; 118pp; English.
The present invention describes a method (M1) for the staged assembly of
a nanostructure using peptide nucleic acids (PNAs). M1 comprises: (a)
contacting a nanostructure intermediate (NSI) having at least one unbound
joining element (JE) with an assembly unit (AU) that comprises several
different JE where: (i) none of these JE can interact with itself or
other JE; and (ii) only one JE in AU and a single unbound JE in NSI are
complementary, so that AU becomes non-covalently linked to NSI to produce
a new NSI for use in subsequent cycles; (b) removing unbound AU; and (c)
cyclic repetition of (a) and (b) to form a nanostructure. The new feature
is that the complementary JE in at least one cycle are PNAs. Also
described are nanostructures formed from many AU, comprising different
JE, where at least one AU includes PNA. M1 is useful for producing
nanostructures with a very wide range of potential applications, e.g.
structural reinforcements (for aerogels, paper, plastics or cement,
particularly as long fibres to improve tensile strength); identification
(anti-counterfeiting) markers; enzyme or catalyst supports; assembly
scaffolds; for construction of nanowires or nanocircuits; size markers
for electron microscopy; molecular sieves and filters; substrates for
optical and other surface coatings; scaffolds for solubilising enzymes or
for trapping, protecting and delivering specific molecules; in high-
density computer memories; as artificial zeolite for absorbing ions from
water and for construction of new materials, including use in biosensors.
PNAs are more homogeneous than inorganic nanoparticles with predictable geometry
form nanostructures, so will produce structures with predictable geometry
and stoichiometry. The present sequence represents a protein containing
coiled coil dimerisation sequences that can be used for structural
elements of assembly units, given in the exemplification of the present
invention.
Sequence 961 AA;
Query Match 7.8%; Score 118; DB 7; Length 961;
Best Local Similarity 22.9%; Pred. No. 0.79;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;
QY 7 DKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRFKOEYSQAASVLVGD 65
Db 635 DKKEBEVKTLFQHDHSIVTHYKNMIREQDLQLEELKQKQISTLKQNEQLQTAVTQVQS 694
QY 66 K-----TLLMDSDQK--YFEATQTVYVCGVATQLLAAYILLPDEYNEKAS-- 110
Db 695 QQHKQYNNLLKVQLGKDSHQHQGPYTDGAQ---MNGVQPEEISR---LREEIEELKSRE 747
QY 111 -----AQKDILIKLVLDGDTK--LNEAKSLVSSQSFNNASGKLALDLSOLTNDPSEK 162
Db 748 LLQSLAELKXDSLIENTLKSSQLSPGTNEQSSATAGDSEQIAELKQELATLKSQ--NSQSV 806

QY 163 SSYFQSOVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEGKL-----IPELKNKJK 216
Db 807 ITKLQTEKQELLQKTEAFAPVGESEVIAKTTD--VEGKLSALLQETKELKNEIK 864
QY 217 SVQNFFTLNTVQKANKIDAAKLTTEIAAIGETKTTETTRFVYVDVDDMLSLKE 276
Db 865 ALSEERTAIKEQLDSSNS-----TIALQNEKNKLEVDITDSKKEQDILLV-LLAD 914
QY 277 AAKMINTCNEYOK 290
Db 915 QDQKIFSLKNKLE 928
RESULT 12
ADL99406
ID ADL99406 standard; protein; 961 AA.
XX
XX ADL99406;
XX
XX 20-MAY-2004 (first entry)
DT Nanostructure assembly protein #57.
DE
XX peptide nucleic acid; PNA; nanostructure.
XX
XX Synthetic.
XX
XX US2003215903-A1.
XX
XX 20-NOV-2003.
XX
XX 21-FEB-2003; 2003US-00370685.
XX
XX 21-FEB-2002; 2002US-00080608.
XX
XX (HYMA/) HYMAN P L.
XX (GOLD/) GOLDBERG E B.
XX
XX Hyman PL, Goldberg EB;
XX
XX WPI; 2004-021840/02.
XX
XX Staged assembly of a nanostructure containing peptide nucleic acid
PT assembly units comprises contacting a nanostructure intermediate with an
PT assembly unit comprising different joining elements, and removing unbound
PT assembly units.
XX
XX Disclosure; Page 43-44; 73pp; English.
XX
XX The invention relates to staged assembly of a nanostructure comprising:
CC (a) contacting a nanostructure intermediate comprising at least one
CC unbound joining element with an assembly unit comprising different
CC joining elements; (b) removing unbound assembly units; and (c) repeating
CC steps (a)-(b) for a sufficient number of cycles to form a nanostructure,
CC where the assembly unit in at least one cycle comprises a peptide nucleic
CC acid. A single joining element of the different joining elements and a
CC single unbound joining element of the nanostructure intermediate are
CC complementary joining elements, where the assembly unit is non-covalently
CC bound to the nanostructure intermediate to form a new nanostructure
CC intermediate for use in subsequent cycles. The method for staged assembly
CC of a nanostructure further comprises: (i) capping the nanostructure with
CC at least one capping unit; and (ii) post-assembly conversion of specific
CC non-covalent interactions of complementary joining elements to covalent
CC linkages, where the linkages are stabilised. The nanostructure
CC intermediate comprises a surface bound initiator assembly unit. A first
CC assembly unit used in at least one cycle comprises at least one
CC structural element covalently linked to a first joining element
CC comprising a peptide nucleic acid, or a first structural element bound to
CC a second structural element to form a stable complex. The structural
CC element is covalently linked to the first joining element and to a second
CC joining element, comprising a peptide nucleic acid. The assembly unit
CC further comprises a functional element, comprising photoactive molecule,
CC photonic nanoparticle, inorganic ion, carbon nanotube, lectin or

CC chemiluminescent molecule. The assembly unit comprises sub-assembly units
CC that bind to each other to form a complex. The present sequence
CC represents a protein containing a coiled coil dimerisation sequence that
CC can be used as a structural element in the method of the invention.
XX
SQ Sequence 961 AA;
Query Match 7.8%; Score 118; DB 8; Length 961;
Best Local Similarity 22.9%; Pred. No. 0.79;
Matches 72; Conservative 54; Mismatches 138; Indels 50; Gaps 12;
QY 7 DKTVEVVKNAIETADGALDLY-NKYLDQVVPWOTFDETIKELSRFKQEYSQAASVVLGDI 65
Db 635 DKKEEVKKTLEQHSISIVHYKNMIREQDLQELKQOISTLKQNEQLQTAVTQVYSQI 694
QY 66 K-----TLMDSDQK--YFEATQTVYEWGCVATQLLAAYILLFDEYNKKAS-- 110
Db 695 QQHKQDQYNLLKVLGKDSQHQGPYTDGAQ---MNGVQPEEISR---LREEIEELKSNRE 747
QY 111 -----AKDILIKVLDDGITK--LNEAQSLLVSSQFNNSGKLLALQSOLTDFSEK 162
Db 748 LLOSQAELKDSLIENLKSSQLSPGTNEQSSATAGDSEQIAELKQELATLKSQ-NSOSVE 806
QY 163 SSYFQSOVDKIRKEAVAGAAAGVAGPGLIISYSIAAGVVEGKL-----IPELKNKJK 216
Db 807 ITKLQTEKQELLQKTEAFAPVGESEVIAKTTD--VEGKLSALLQETKELKNEIK 864
QY 217 SVQNFFTLNTVQKANKIDAAKLTTEIAAIGETKTTETTRFVYVDVDDMLSLKE 276
Db 865 ALSEERTAIKEQLDSSNS-----TIALQNEKNKLEVDITDSKKEQDILLV-LLAD 914
QY 277 AAKMINTCNEYOK 290
Db 915 QDQKIFSLKNKLE 928
RESULT 13
ABP39975
ID ABP39975 standard; protein; 1211 AA.
XX
XX ABP39975;
XX
XX 24-JUL-2002 (first entry)
DT
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4820.
DE
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-00134001.
XX
XX 14-AUG-1997; 97US-0055779P.
XX 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN92520.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
XX Disclosure; SEQ ID NO 4820; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading

CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences can
 CC also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life cycle
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from the USPTO web site
 XX
 XX Sequence 1211 AA;

Query Match 7.8%; Score 118; DB 5; Length 1211;
 Best Local Similarity 19.3%; Pred. No. 1.1;
 Matches 69; Conservative 53; Mismatches 112; Indels 124; Gaps 13;
 QY 5 VADKTVVVK-----NAIETADGALDLYNKYLDVPIPWOTFDTIKELGRFKQVSOAA 58
 Db 183 IIEESAGVLKVKRKAESIQKLDHTEDNLNRVEDILYDLGRVPLKEEAAIAKEYKQLS 242
 QY 59 S-----VLVDIKTLMDSD-----KYFEATQTVYVCGVATQLLAAYILLFDEYN 105
 Db 243 KMEQSDVITVSDIDHTEDNQRLDERLNHLKSQAQKE--GQAQINQ-----LQKYK 296
 QY 106 EKKASAKDILIKVLDDGITKNEAKSLLVSSQSFNNASGKLALDLSOLTND-----F 159
 Db 297 GKRQNDYD-----IEKLN-----YELVKATENYQSLGNVLEERKKNQSETWARY 344
 QY 160 SEKSSYFQSQVDKIRKEAVAGAAAGVAGPFLGLIISYIAAGVVEGKLIPELKN----- 213
 Db 345 EELNDLESQIDSIKNEK-----AKNEKLLADLKNKQKOLN 380
 QY 214 -----KLKSVQNFVTLNTVQANKDT-----DAKLLKT 244
 Db 381 KEVOELESLLYISDEQHEKLEIKNSYYTLMSQSVVNDIRFLEHTINEAKKSLD 440
 QY 245 TEATAAGEIKTETTRFYVDVDDMLSLLEAKKQWINTCNEYQKRHGKTKLFFVPE 302
 Db 441 SRLVE-----AFNQLKDIQQNITQTOKEYQS--SKKSMKEVEQ 476

RESULT 14
 AAR43563
 ID AAR43563 standard; protein; 476 AA.
 XX
 AC AAR43563;
 XX
 XX 25-MAR-2003 (revised)
 DT 09-JAN-2003 (revised)
 DT 05-APR-1994 (first entry)
 XX
 DE Hyaluronan receptor.
 XX
 XX Hyaluronan binding protein; HA; RHAMM; mediated motility; wound; healing;
 KW diagnosis; treatment; cell locomotion; tumour invasion; birth defects;
 KW inflammatory disorder; Alzheimer's disease; dementia;
 KW Parkinson's diseases; Huntington's disease; AIDS; diabetes; auto;
 KW immune diseases; corneal dysplasia; hypertrophy; surgery; burns; strokes;
 KW multiple sclerosis; depression; schizophrenia; CNJ; contraception;
 KW in vitro fertilisation; embryo development.
 XX
 OS Rattus sp.
 XX
 PN WO9321312-A1.
 XX
 PD 28-OCT-1993.
 XX
 PF 13-APR-1993; 93WO-CA000158.
 XX
 PR 09-APR-1992; 92GB-00007949.
 XX
 XX (UYMA-) UNIV MANITOBA.
 PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.

XX Turley EA;
 XX
 XX WPI; 1993-351722/44.
 DR N-PSDB; NAO51212.
 XX
 PT DNA encoding hyaluronan receptor - used to produce proteins and
 PT antibodies for alteration of cell locomotion.
 XX
 PS Claim 7; Fig 23; 88pp; English.
 XX
 XX The sequence is that encoded by a cDNA clone encoding the hyaluronan
 CC receptor (HARC). The sequence was obtd. by screening a 3T3 library in
 CC lambda gt11 with antibodies to HARC. A clone of 1.9 kb was obtained and
 CC used to rescreen the library to obtain the full length, 2.9 kb clone. HA
 CC is down regulated in stationary normal cells and is only expressed in
 CC situations where cell motility is desired, e.g. in wound healing, in
 CC response to growth factors and in chemotaxis by white blood cells. HA may
 CC be used for diagnosis and treatment of diseases involving cell
 CC locomotion, e.g. tumour invasion, birth defects, acute and chronic
 CC inflammatory disorders, Alzheimer's and other forms of dementia, AIDS,
 CC diabetes, autoimmune diseases, corneal dysplasias and hypertrophies,
 CC burns, surgical incisions and adhesions, strokes, multiple sclerosis,
 CC depression/schizophrenia related to neuronal growth and pain states
 CC involving nerve sprouting; also in CNJ and spinal cord regeneration,
 CC contraception, in vitro fertilisation and embryo development. See also
 CC AAR46548-51. (Updated on 09-JAN-2003 to add missing OS field.) (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 476 AA;

Query Match 7.8%; Score 117.5; DB 2; Length 476;
 Best Local Similarity 20.8%; Pred. No. 0.33;
 Matches 69; Conservative 73; Mismatches 125; Indels 65; Gaps 18;
 QY 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDVPIPWOTFDTIKELSRFKQVSOASV 60
 Db 99 LDNLLREKEVELEKHIAQAAILIAQEKYIDTA---QSLRVVTAQLESVQKYNDAQS 155
 QY 61 LVGDIKTLMDSDQKYFEATQTVYVCGVATQLLAAYILLFDEYNEKKAQAQDILIKVL 120
 Db 156 L-RDVTQAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 205
 QY 121 DDGITKLEAKQSLL-VSSQSFNNASGKLAL-DSQLTN-DFSEKSSYFQSQVDKIRKEA 177
 Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLLENLTLOEKVAMAEKSVEDVQQOI 265
 QY 178 YAGAAAGVAGPFLGLIISYIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
 Db 266 LTAESTNQ-----EYA-----RMVDLQNRSTLKEEIKETSSFLEKITDLK 308
 QY 227 NTVKQANKDI-----DAAKLLKTETI-----AAIGKTKTETTRFYVD-YDD 268
 Db 309 NQLRQDEDFRQLEKGRKTAENVNVTETMTMEINKWALLYELYEKTQPFQQLDAFEA 368
 QY 269 LMLSLLKE--AAKQWINTCNE-YQKRHGKTKL 297
 Db 369 EKQALLNEHGATQEQELNKIRDSYAQLLGHQNL 400

RESULT 15
 ABB49720
 ID ABB49720 standard; protein; 927 AA.
 XX
 AC ABB49720;
 XX
 XX 05-FEB-2002 (first entry)
 DT
 DS Listeria monocytogenes protein #2424.
 XX
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX

OS Listeria monocytogenes.
 PN WO200177335-A2.
 XX 18-OCT-2001.
 XX 11-APR-2001; 2001WO-FR001118.
 XX 11-APR-2000; 2000FR-00004629.
 XX (INSP) INST PASTEUR.
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX WPI; 2002-010914/01.
 DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and related
 PT polypeptides.
 XX Claim 6; SEQ ID NO 2425; 192pp; French.
 PS The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 927 AA;
 SQ
 Query Match 7.8%; Score 117.5; DB 5; Length 927;
 Best Local Similarity 21.1%; Pred. No. 0.83;
 Matches 73; Conservative 33; Mismatches 141; Indels 99; Gaps 11;
 QY 3 EIVADKTVVVKNAETADGALDLYNKVLDVDPWOTFDETI-----KELSRFKQVYSQAA 58
 DB 166 EATPDKIKESGDGPAADGS-----GKIDGLVKVKSQGNKTTISTNLKTLADSSLTFPKDGA 221
 QY 59 SVL-----VGDIK-----TLLMDSQDKYF 77
 DB 222 NTLEVLGKTYTDGNTAAAGDKLKNAGVSTLAAGVGLKDGVAALDGGATKLASGVSTYT 281
 QY 78 EATQTVYEWCVAPQLLAAYILLFDEYNEKKAQKDILIKVLDGDTIKLNEAQSLLVS 137
 DB 282 SGVDTL---AGGINQAVTGSTALSDGLNKNMS-----VPTLASGITQLNNGQKSLATG 332
 QY 138 SQSFNNAAGKLLA-----LDSQLTN-----DFSEKSSYFQSDVKIRKAYAGAAAGVA 187
 DB 333 LDSLVDSGNKLSAGLXELDGNLTDKQCKIAQLKQGMNDLQOQIDQLNQSVNGEDAA---- 388
 QY 188 GPFGLIISYSIAAGVVEGKILPELKNKLSQNFVFTLSNTVKQANKDIDAALKLTTEI 247
 DB 389 -----LAKQLATLQKLSLSDLQNLGTLTIKS-----NANFDAEAIKSKINATA 429

QY 248 AAI GEIKTETETTRFYVDYDDMLSLKEAAKQKQKNTCNEYQKRHG 293
 DB 430 GVS AEDKQKI-----IDAIQALDKETQKATQVATVEQLQSG 467

RESULT 16
 AAG82283
 ID AAG82283 standard; protein; 885 AA.
 XX AC AAG82283;
 XX DT 03-SEP-2001 (first entry)
 XX DS S. epidermidis open reading frame protein sequence SEQ ID NO:1660.
 XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
 XX endocarditis.
 XX OS Staphylococcus epidermidis.
 XX PN WO200134809-A2.
 XX PD 17-MAY-2001.
 XX PF 09-NOV-2000; 2000WO-US030782.
 XX PR 09-NOV-1999; 99US-0164258P.
 XX (GLAX) GLAXO GROUP LTD.
 XX PI Kimmerly WJ;
 XX DR WPI; 2001-316495/33.
 XX N-PSDB; AAH53133.
 XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis.
 XX Claim 18; Page 462; 2188pp; English.
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
 CC and (II) can have antibacterial activity and therefore can be used in
 CC vaccination. The nucleic acids (I) may be used to produce the S.
 CC epidermidis polypeptides (II) via the production of vectors containing
 CC them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
 CC represent oligonucleotide sequences and primers which are used in the
 CC exemplification of the present invention. N.B. The present invention
 CC specifically claims all the polynucleotide sequences given in the
 CC sequence listing of the present specification, however the sequence
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
 CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 CC for SEQ ID NO:4455 to 4464
 XX SQ Sequence 885 AA;
 Query Match 7.7%; Score 116; DB 4; Length 885;
 Best Local Similarity 20.0%; Pred. No. 1;
 Matches 63; Conservative 42; Mismatches 92; Indels 118; Gaps 12;
 QY 42 ETIKELSRFKQVYSQAA-----VLVGDITKLLMDSQD-----KYFRATQTVYEWCG 88
 DB 2 EPLKEEAIAKEVYKQLSKEMEQQSDVIVTSDIDHYEDNQRLDLRLNHLKSQAQKE--G 59
 QY 89 VATQLLAAYILLFDEYNEKKAQKDILIKVLDGDTIKLNEAQSLLVSQSFNNAAGK 148

QY 6 ADKTVVVKNALETADGALDLYNNKYLDQVIPWQTFDETIKELSRFKQE-----YSQA 5

Db 103 ATEKLEVAKEATQ-----DKVEKTSQSLVEDIKDKAQSLOEDAADTVLEALKQA 149
Qy 58 ASVLVGDIKTLMLMSQDKYFEATQTVYEWG-----GVATQLLAAYILLFDEYN--- 105
Db 150 AS---DKVETTKABAQSLKDDATQT-FESAQAQVEGKVEAIKEQVLQVDSLKDDTTQDN 205
Qy 106 ---EKKASAKDIIILIKVLDDGITKLNQAQSLLYSSQSP-NNASGKL-----LALD--SQ 154
Db 206 TDQDEQKTLKQKAAQATAAKRKVEDVDVKGHTESFRNTAGSKIDEIKQAQAVDKTEE 265
Qy 155 LTNDSEKSSYFQSDVKIRKEAYAGAAAGVVGPFGLIISYSTAAAGVVEGKLIPELKNK 214
Db 266 VKSLSQKADALKSSGSELKQTAQT-AANDAITAQAQVSGSVAAADSQAQSTAQSAKDX 324
Qy 215 LKSVQNFPTLSNTVQKQNDIDAQKLTTEIAAIGIKETETETTRFYDDYDMLSL 274
Db 325 L-----NQLFEQKSGALDEKQVQELGKFKGATEKINAVSENVDLATQVI 367
Qy 275 KEAKKMWNTC-----NEVQKRGHKKTL 297
Db 368 KEEAQAQLTNAQBSLQAQKAAGEBYDATHDKGL 401

RESULT 19

ABBS9344

ID ABB59344 standard; protein; 2056 AA.

XX AC ABB59344;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 4824.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-0061415P.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03447.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signaling and cell-cell

XX PT interactions.

XX PS Disclosure; SEQ ID NO 4824; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-

XX CC ABBS7072). The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX CC Sequence 2056 AA;

Query Match 7.68; Score 115; DB 4; Length 2056;
Best Local Similarity 21.3%; Pred. No. 4;
Matches 80; Conservative 48; Mismatches 107; Indels 140; Gaps 16;
Qy 14 KNALETADGALDLYNKYL-----DQVIPWOTFDETIKELSRFK 51
Db 1302 KTVLEKAKGTLEAENADLATELSVNSSRQENDRRRKQAESQIAELQV---KLAIEERAR 1358
Qy 52 QEYSQAASVLVGDITKTLMLMSQDKYFEATQTVYEWGCVATQLLAAYILLFDEYNK----- 107
Db 1359 SELQEKCTKLQEAENITNQLAEALKASAAVKSASNMESQLTEAQQLLBEETRQKLGSL 1418
Qy 108 -----KASAQKDI-LIKVLDDGIT 125
Db 1419 SKLRQISEKALQEQLEEDDEAKRNYERKLAEVTTQMQEIKKKAEBDADLAKELEGKK 1478
Qy 126 KLN-----BAQ-KSLLVSGSFNNASGKL-----LALDSQLTNDFS-EKSYFQSQ 169
Db 1479 RLNKDIEALERQVKELIAQNDRLDKKKIQSELEDATIEAORTKVLELEKK---QKN 1535
Qy 170 VDKIRKEAYAGAAAGVVGPFGLIISYSIA-----AGVVEGKL----- 207
Db 1536 FDKILAIEKA-----ISQIAQERDTABREAREKETKVLVSRELDEAFDK 1581
Qy 208 IPELKNKLKSVQNFPTLSNTVQKQNK---DIDAQKLTTEIAAIGIKETETETTRFYV 264
Db 1582 IEDLENKRTQLNELDDLANTQGTADKNVHELEKAKKALESQLA---ELKAQNEELE--- 1635
Qy 265 DYDDMLSLLEKAAK 279
Db 1636 --DDLQLT---EDAK 1645

RESULT 20

ABM68103

ID ABM68103 standard; protein; 582 AA.

XX AC ABM68103;

XX DT 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #1200.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

XX KW detection; food; gene expression; plant; animal; microorganism; toxin;

XX KW antibiotic; biopesticide; virulence factor; disease model; plague;

XX KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN W0200294867-A2.

XX PD 28-NOV-2002.

XX PF 07-FEB-2002; 2002WO-IB003040.

XX PR 07-FEB-2001; 2001FR-00001659.

XX PA (INSP) INST PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

XX PI Buchrieser C;

XX DR WPI; 2003-148459/14.

XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,

XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 1200; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded

XX CC proteins from Photorhabdus luminescens. The isolated sequences are

	Query Match	7.6%	Score 114.5;	DB 6;	Length 1189;	
	Best Local Similarity	20.1%;	Pred. No. 2.1;			
	Matches 67;	Conservative 45;	Mismatches 96;	Indels 125;	Gaps 13;	
QY	5 VADKTVEVK-----NAIETADGALDLYNKLDQVIPWQTFDTIKELSRFKOYSQA 58	: : : :	: : : :	: : : :	: : : :	
Dd	161 IIESAGVLKYKKAESIOKLHTEDLNLRVEDILDLEGRVPLKEAAIAKEYKOLS 220	: : : :	: : : :	: : : :	: : : :	
QY	59 S-----VLVGDIKTLMLDSQKYFEATQTVVEMC-----GVATQLLAA 96	: : : :	: : : :	: : : :	: : : :	
Dd	221 KEMEQSVDIVTVSDI-----DHYTEDNRQRLDERLNHLKSQAQEKGGQAQINQLQR 272	: : : :	: : : :	: : : :	: : : :	
QY	97 YILLFDEYNKKASQAOKDILIKVLDDGTIKLNEAQSLLVSSQSFNNASGKLLALDSQIT 156	: : : :	: : : :	: : : :	: : : :	
Dd	273 Y-----KGKRQQN-----DYDEKLN-----VELVKATENYEQISGLKNLVLEERKK 313	: : : :	: : : :	: : : :	: : : :	
QY	157 ND-----PSEKSSVFOSQVDKIRKAYAGAAGVVGPFGLIIISYSIAAGVVEGKIPIE 210	: : : :	: : : :	: : : :	: : : :	
Dd	314 NOSETNARYEBELDNLSQIOISIKNEK-----AQNEKLLAD 349	: : : :	: : : :	: : : :	: : : :	
QY	211 LKN-----KLKSVQNFFTLTSLNTVKQANKDI----- 236	: : : :	: : : :	: : : :	: : : :	


```
PS PS 194 ISYSIAAGVVEGKLIPELKNKLSQVNFETTLGNTVKQA-----NKDIDAAKLKLTTETI 247
XX XX 312 -----KHTQKSSQEQYQISLAERAKQKVKHFEKQLQESLMAAAEKETEI 353
CC CC 248 --AAIGBIKTETE-----TTRFYVDYDLMISLLKEAKKQMINTCNEYQKEH----- 292
CC CC 354 QKAEANLIKQTELEKQKSTKELLAEIRQYVDLMQEQA--AVGNELKYLERQYIQETA 411
CC CC 293 -GKTKLFVPEPV 303
CC CC 412 KSKOTLAKQSEV 423

RESULT 27
AAR99673
ID AAR99673 standard; protein; 606 AA.
XX
AC AAR99673;
XX
XX 10-OCT-1996 (first entry)
XX
DE Receptor for hyaluronic acid mediated motility RHAMM 1.
XX
XX RHAMM 1; receptor for hyaluronic acid mediated motility;
KW hyaluronan receptor; cell locomotion; cell proliferation; breast cancer;
KW therapy.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
FH Modified-site 91..93 /label= N-glycosylation_site
FT FT Modified-site 258..260 /label= N-glycosylation_site
FT FT Modified-site 279..281 /label= N-glycosylation_site
FT FT Modified-site 300..302 /label= N-glycosylation_site
FT FT Modified-site 321..323 /label= N-glycosylation_site
FT FT Modified-site 342..344 /label= N-glycosylation_site
FT FT Modified-site 373..375 /label= N-glycosylation_site
FT FT Modified-site 413..415 /label= N-glycosylation_site
FT FT Domain 532..542 /label= Hyaluronan_binding_domain-I
FT FT Domain 553..562 /label= Hyaluronan_binding_domain-I
FT FT Modified-site 594..596 /label= N-glycosylation_site
XX XX EP721012-A2.
XX PN 10-JUL-1996.
XX PD 16-OCT-1995; 95EP-00307310.
XX XX 14-OCT-1994; 94GB-00020740.
XX XX (UYMA-) UNIV MANITOBA.
XX PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
XX PI Turley EA, Zhang S, Entwistle J;
XX
XX WPI; 1996-310997/32.
DR N-PSDB; AAT34499, AAT34500, AAT34524, AAT23423.
XX
XX Receptor for hyaluronic acid-mediated motility protein, and DNA encoding
PT it - useful to treat or prevent diseases associated with the receptor,
PT e.g. breast cancer.
XX
```

```
PS PS Claim 2; Page 40-42; 117pp; English.
XX
CC RHAMM 1 (AAR99673), or Receptor for Hyaluronic Acid Mediated Motility, is
CC a hyaluronan receptor protein which is involved in cell locomotion or
CC motility and cell proliferation and transformation. Its amino acid
CC sequence was deduced from a cDNA clone (AAT34499) obt'd. from murine 3T3
CC cells and a genomic sequence (AAT34500) from a mouse fibroblast genomic
CC library. 2 Alternative mRNAs for RHAMM 1 were found, RHAMM 1A (see also
CC AAT34523) and RHAMM 1B (AAT34524), that had identical translated
CC portions. Increased expression of RHAMM 1 protein is indicative of a poor
CC prognosis for breast cancer. The protein can be used to suppress or
CC control a tumour by modulating the interaction of cell-associated RHAMM
CC with its ligand
XX
XX Sequence 606 AA;
XX
XX Query Match 7.4%; Score 112; DB 2; Length 606;
XX Best Local Similarity 20.5%; Pred. No. 1.4;
XX Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;
XX
XX 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETTELKELSRFKQEYSQASV 60
Db 228 LONLLREKEVELEKHAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVEKQKYNDRQS 284
XX
XX 61 LVGDIKTLMSQDKYFEATQTVYEWCGVATOLLAAVILLFDEYNEKKAQKDLIKVL 120
Db 285 L-RDVTVAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQSRLRDVTAQ-L 334
XX
XX 121 DDGITTCLNEAQAQKSL--VSSQSFNNAAGKLAL--DSQLTN--DFSEKSSYFQSOVDKIRKEA 177
Db 335 ESVOEKYNDTAQSLRDVTAQLESYKSTLKEIEDKLENLTLOEKVAAKESVEDVQQOI 394
XX
XX 178 YAGAAAGVWAGPGLIISYIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
Db 395 LTAESTNQ-----EVA-----RWQDLQNRSTLKEEIKETSSFLEKITDLK 437
XX
XX 227 NTVKQANKOI-----DAAKLLKLTETIAAIGSIKTTETTRFYVDYDDL 269
Db 438 NQLRQDEDFRQLEKKGKETAENVMTELT-----MEINKWRLLYDEL 482
XX
XX RESULT 28
XX AAR99675
ID AAR99675 standard; protein; 631 AA.
XX
XX AC AAR99675;
XX
XX 10-OCT-1996 (first entry)
XX
XX RHAMM 1-2a isoform.
XX
XX RHAMM 1-2a; receptor for hyaluronic acid mediated motility;
KW hyaluronan receptor; cell locomotion; cell proliferation; breast cancer;
KW therapy.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
FH Key 55..79
FT Region /note= "exon 2A-encoded region"
XX
XX EP721012-A2.
XX
XX 10-JUL-1996.
XX
XX 16-OCT-1995; 95EP-00307310.
XX
XX 14-OCT-1994; 94GB-00020740.
XX
XX (UYMA-) UNIV MANITOBA.
XX PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
XX PI Turley EA, Zhang S, Entwistle J;
```


PD	27-AUG-2003.	ADP04710	ID	ADP04710 standard; protein; 808 AA.
XX		XX	AC	ADP04710;
PF	20-DEC-2002; 2002EP-00102902.	XX	AD	ADP04710;
PR	20-DEC-2001; 2001EP-00130253.	XX	DT	12-FEB-2004 (first entry)
XX	(CELL-) CELLZOME AG.	XX	DE	Bacterial polypeptide #823.
XX	Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;	XX	XX	Proteus mirabilis infection; bacterial infection; antibacterial;
PI	Marzoch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;	KW	KW	immunostimulant.
PI	Michon A, Leutwein C, Rick J;	XX	OS	Proteus mirabilis.
XX	WPI; 2003-638460/61.	XX	PN	US6605709-B1.
DR	N-PSDB; ADK64381.	XX	XX	12-AUG-2003.
XX	New proteins and protein complexes from eukaryotes, useful as targets in	PD	XX	05-APR-2000; 2000US-00543681.
PT	drug screening, or in diagnosing or screening for the presence of a	PF	XX	09-APR-1999; 99US-0128706P.
PT	disease or disorder, or a predisposition for developing a disease or	PR	XX	(GENO-) GENOME THERAPEUTICS CORP.
PT	disorder in a subject.	PA	XX	Breton GL;
XX	Disclosure: SEQ ID NO 2369; 13pp; English.	PI	XX	WPI; 2003-895291/82.
XX	The invention relates to novel protein complexes comprising a first and a	DR	XX	N-PSDB; ADP00538.
XX	second protein, or its derivative, fragment, homologue or variant. The	XX	XX	New Proteus mirabilis polypeptides and polynucleotides, useful as
CC	proteins are selected from given protein complexes, which are not defined	PT	PT	reagents for diagnosis of bacterial disease, as components of
CC	in the specification. The variants are encoded by nucleic acids that	PT	PT	antibacterial vaccines, as targets for antibacterial drugs, or as
CC	hybridize to the nucleic acids encoding the proteins under low stringency	PT	PT	biocontrol agents for plants.
CC	conditions. The protein complexes are useful as targets for an active	XX	XX	Disclosure: SEQ ID NO 4995; 870pp; English.
CC	agent of a pharmaceutical. These protein complexes are particularly	XX	XX	The invention relates to new Proteus mirabilis polypeptides and
CC	useful as drug targets for the treatment or preventing of a disease or	CC	CC	polynucleotides. The invention also relates to antibodies against the
CC	disorder. The complexes and methods above are useful in diagnosing or	CC	CC	polypeptides, methods for producing the polypeptides, a method of
CC	screening for the presence of a disease or disorder or a predisposition	CC	CC	generating vaccines for immunising an individual against P. mirabilis, a
CC	for developing a disease or disorder in a subject. These are also useful	CC	CC	method for evaluating a compound for the ability to bind a P. mirabilis
CC	in screening for a drug for treatment or prevention of a disease or	CC	CC	polypeptide and a method for screening test compounds for anti-bacterial
CC	disorder. The molecule that modulates the amount, activity or protein	CC	CC	activity. The polypeptides and polynucleotides are useful as molecular
CC	components of the complex is useful for the manufacture of a medicament	CC	CC	targets for diagnosing, preventing and treating pathological conditions
CC	for the treatment or prevention of a disease or disorder. This sequence	CC	CC	resulting from bacterial infection, as reagents for diagnosis of
CC	corresponds to a protein of the invention. (Note: the sequence data for	CC	CC	bacterial diseases, as components of antibacterial vaccines, as targets
CC	this patent did not form part of the printed specification but was	CC	CC	for antibacterial drugs or as bio-control agents for plants. This
CC	obtained from the EPO in electronic format).	CC	CC	sequence represents a Proteus mirabilis polypeptide of the invention.
XX	Sequence 1875 AA;	XX	SQ	Sequence 808 AA;
XX	Query Match 7.4%; Score 112; DB 7; Length 1875;	XX	QY	Query Match 7.4%; Score 111.5; DB 7; Length 808;
XX	Best Local Similarity 19.7%; Pred. No. 6.3;	XX	Db	Best Local Similarity 22.0%; Pred. No. 2.2;
XX	Matches 68; Conservative 66; Mismatches 122; Indels 90; Gaps 15;	XX	QY	Matches 66; Conservative 55; Mismatches 108; Indels 71; Gaps 14;
QY	2 TEIVADKTVVVKNAIETADGALDLYNK---VLDQVI-----PWQTFDETIK----- 45	QY	4	IVADKTVVVKNAIETADGALDLYNKYLQDQVPMQTFDETIKELSRPKQZ-YSOAAASVLV 62
Db	685 SRLAEERFKLLSNTLDTLTKAENDQLRKRFYDLQNTILKQDSKTHETLNEYVCSCKLSI 744	Db	257	IIIESEAKLVKKIQTSETIQELNNAI--LASNV---RTGRELHKYKQKCAFCGNILP 311
QY	46 ---ELSRFKQZ-----SQAAASVLVGDIKTLLMDSODKYFEATQT 82	QY	63	GDITKLLMDSODKYFEATQTVVWCGVATQLLAAYILLDFEYNEKKASAKQDILIKVLDD 122
Db	745 VETELNLKEQKRLVHLEKNLQKLNLSPEKDSLRINVLTQTLQKEREDELETRKS 804	Db	312	ED---LLQKIDKHNKES-----ELHDE-----SLNILLSTIEK 342
QY	83 VYEWCGVATQLLAAYILLDFEYNE-----KKASAKQDILIKVL--DGIKTLNEAKSLIV 136	QY	123	GITKTLNEAKSLIVSSQSFNNASGKLALD---SQTNDSEKSSYFOSQVDKTRKAYA 179
Db	805 CQK-----KIDELDALSELKKTSSOKDHIKQLEDNNSNIWYQNKIEA 850	Db	343	ELLKIPNLK--IKTIDFYSNFANDLNTLNDLSKLSIDYSKLSLIKEQIEKRSDFIT 400
QY	137 SSQSFNNASGKLALDLSQLTN---DFSEKSSYFOSQVDKIRKAYAGAAAGVAGPFG 192	QY	180	GAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNLKSQVQFFTLST--VKQAKDID 237
Db	851 LKQDYSEV---ITSVDGSKQTDIEKLQYKVSLEKEIEEDKIRLHTY-----NVMW---ET 899	Db	401	-----PLEFHSVSVSDSINELRYSYE---KIRKSNDFTKSLNTEQVKARN---- 444
QY	193 IISYSIAAGVVEGKL-IPELKNLKSQVQFFTLSTNTVQKANDIDAAKLTETAAIG 251	QY	252	EIKTETETTFYVDYDMLSLLEAKAKMINTCNE--YQKRGKK 295
Db	900 INDSLSLKELEKSKINLTDAVYQIKYKDYETTSQSLQNTSKLDESFDFTNQIKLNT 959	Db	960	DEKTSLED-----KISLLKE-----QMFNLNLELDLQKGNK 992
QY	252 EIKTETETTFYVDYDMLSLLEAKAKMINTCNE--YQKRGKK 295	QY	236	AAKLKLTETIAAIGIKETETTFYVDYDMLSLLEK---AAKMTINTCNEYQKRGK 294
Db	960 DEKTSLED-----KISLLKE-----QMFNLNLELDLQKGNK 992	QY		
XX	RESULT 31	XX		

445 --ELRLHEVYKFTIDIK-----YVDERTVINKLKEKEDLEKTSINTKQNVDEKRKK 49

Db

RESULT 32

ABR57002

ID ABR57002 standard; protein; 435 AA.

XX ABR57002;

XX

DT 05-AUG-2003 (first entry)

XX

DE Mouse RHAMM V-3 region amino acid sequence SEQ ID NO:80.

XX

KW RHAMM; receptor hyaluronic acid mediated motility; hyaluronic acid; HA; neuroprotective; vasotropic; antidiabetic; nootropic; antiarthritic; antiParkinsonian; antiinflammatory; osteopathic; dermatological; vaccine; antiarteriosclerotic; cyostatic; vulnerable; hyaluronic acid antagonist; HA binding inhibitor; multiple sclerosis; restenosis; diabetes mellitus; neurological disease; Alzheimer's disease; Parkinson's disease; cancer; arthritis; inflammatory joint disease; osteoporosis; atherosclerosis; inflammatory dermatosis; inflammatory bowel disease; metastasis; inflammatory disease; respiratory distress syndrome; wound healing; RHAMM binding protein; RABP.

XX

OS Mus sp.

XX

PN WO2003033535-A2.

XX

PD 24-APR-2003.

XX

PF 15-OCT-2002; 2002WO-CA001563.

XX

PP 15-OCT-2001; 2001US-00978309.

PR

PR (TRAN-) TRANSITION THERAPEUTICS INC.

PA

PI Cruz T, Pastrak A, Turley EA;

XX

XX WPI; 2003-372323/35.

XX

XX New P16, P-16d, human and murine P32, S3, S7, and V2 polypeptides, useful for treating e.g. multiple sclerosis, restenosis, diabetes mellitus, Alzheimer's disease, Parkinson's disease, arthritis, cancer and metastasis.

PT

PT Claim 4; Page 28-29; 215pp; English.

PS

XX

CC The present invention describes a polypeptide (I) which binds to hyaluronic acid (HA) and thereby inhibit the binding of HA to RHAMM (receptor hyaluronic acid mediated motility), comprising an amino acid sequence of P16, P-16d, human P32, murine P32, murine S3, human S3, murine S7, human S7, murine V2 and human V2. Also described: (1) a pharmaceutical composition comprising (1); (2) an antibody which binds to (1) or to a polypeptide comprising murine V3 or human V3 amino acid sequence; (3) a vaccine composition for the treatment of multiple sclerosis, restenosis or diabetes mellitus, comprising an antigen for the antibody of (2); and (4) a method for treating multiple sclerosis, restenosis or diabetes mellitus, by administering (1), human or murine V3, or an antibody to these polypeptides. (1) has neuroprotective, vasotropic, antidiabetic, nootropic, antiParkinsonian, antiarthritic, antiinflammatory, osteopathic, dermatological, antiarteriosclerotic, cyostatic and vulnerary activities, and can be used in vaccines and as an HA antagonist. The polypeptides and the antibodies to these polypeptides are useful for treating multiple sclerosis, restenosis or diabetes mellitus, as well as neurological diseases such as Alzheimer's disease, Parkinson's disease, arthritis and other inflammatory joint diseases, osteoporosis, inflammatory dermatosis, inflammatory bowel disease and other inflammatory diseases, atherosclerosis and related diseases, cancer and metastasis, and chronic and acute respiratory distress syndrome. The polypeptides may also be used to promote wound healing. ABR56968 to ABR57022 and ACC79533 to ACC79537 represent sequences used in the exemplification of the present invention

XX

CC antiproliferative or antiinflammatory compound. The methyl donor compound
CC acts synergistically with the antiviral, antiproliferative or
CC antiinflammatory compound to provide an enhanced therapeutic effect with
CC minimal side effects. In the antiinflammatory compositions of the
CC invention, the amount and the frequency of administration is lower than
CC that required for the antiinflammatory compound in the absence of the
CC methyl donor compound. Compositions of the invention can be used in the
CC treatment of viral disease, proliferative diseases (e.g., cancer), and
CC inflammatory diseases such as multiple sclerosis. The present sequence
CC represents a peptide fragment of murine RHAMM (receptor for hyaluronic
CC acid mediated motility) which is specifically claimed as the
CC antiinflammatory compound in an antiinflammatory composition of the
CC invention for the treatment of multiple sclerosis
XX
SQ Sequence 435 AA;
Query Match 7.4%; Score 111; DB 6; Length 435;
Best Local Similarity 20.5%; Pred. No. 1.1;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;
QY 1 MTEIVADKTVVKNVNAITAGALDLYNKYLDQVTPWQTFDETIKELSRFKQESQASV 60
Db 99 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAYILLDFEYNEKKAQAQDILIKVL 120
Db 156 L-RDVTQAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQSRLDVTQA-L 205
QY 121 DDGITKLENAQKSL-L-VSSQSFNNASGKLAL-DSQLTN-DPSEKSSYFQSDVKIRKEA 177
Db 206 ESQVEKYNDTAQSLRDVSAQLESYKSTLKEIEDLLENLTLOEKVAMAEKSVEDVQQOI 265
QY 178 YAGAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIKETSSFLEKITDLK 308
QY 227 NTVQKANKDI-----DAKLLKLTTEIAAIGEIKETETTRFFVVDYDDL 269
Db 309 NQLRQODEDFRKLEEGKRTAEKENVMTELT-----MEINKWRLLYEEL 353
RESULT 34
ADCO2471
ID ADC02471 standard; protein; 435 AA.
XX
AC ADC02471;
XX
XX 18-DEC-2003 (first entry)
XX
XX Mouse RHAMM V3 protein.
XX
XX RHAMM; Receptor hyaluronic acid mediated motility;
XX cell surface target protein; AP-1; activating protein-1;
XX response-to-injury; vaccine; inflammatory neurological disorder;
XX diabetes mellitus; arthritis; inflammatory dermatosis;
XX inflammatory bowel disease; cancer; kidney fibrosis;
XX inflammatory lung disease; obesity; lupus; cardiovascular disease; wound;
XX multiple sclerosis; Parkinson's disease; Alzheimer's disease;
XX alpha helical peptide; mouse.
XX
OS Mus sp.
XX
XX US2003100490-A1.
XX
XX 29-MAY-2003.
XX
XX 15-OCT-2001; 2001US-00978309.
XX
XX 01-APR-1999; 99US-0127457P.
XX
XX 03-APR-2000; 2000US-00541522.
XX
XX 05-OCT-2000; 2000US-00685010.
XX
XX (CRUZ/) CRUZ T.

(PAST/) PASTRAK A.
(TUR/) TURLEY E A.
Cruz T, Pastrak A, Turley EA;
WPI; 2003-755209/71.
New polypeptide for diagnosing, preventing or treating inflammatory
neurological disorders, cancer, obesity, lupus or diabetes mellitus, S7
comprises an amino acid sequence of P16, P32, or murine or human S3, S7
or V2.
Disclosure; SEQ ID NO 80; 110pp; English.
XX
XX The invention relates to a polypeptide comprising an amino acid sequence
XX selected from the sequences of P16, P32, murine S3, human S3, murine S7,
XX human S7, murine V2 and human V2, all derived from the protein RHAMM
XX (receptor hyaluronic acid mediated motility) a cell surface target
XX protein that is required for activation of the AP-1 (activating protein-
XX 1) pathway on response to injury. Also included are an antibody that
XX binds to the polypeptide cited above, a pharmaceutical or vaccine
XX compositions for treating an inflammatory neurological disorder or
XX for the antibody cited above) and treating an inflammatory neurological
XX diabetes mellitus (comprising the above amino acid sequence or an antigen
XX disorder, arthritis, inflammatory dermatosis, inflammatory bowel disease,
XX cancer, kidney fibrosis, inflammatory lung disease, obesity, lupus,
XX cardiovascular disease, diabetes mellitus or wounds (comprising
XX administering to a patient the above polypeptide or antibody. In the
XX above method, the inflammatory neurological disorder is multiple
XX sclerosis, Parkinson's disease or Alzheimer's disease. The polypeptide or
XX antibody may also be used for diagnosing or preventing the above
XX diseases. The present sequence is a mouse RHAMM active protein fragment.
XX
SQ Sequence 435 AA;
Query Match 7.4%; Score 111; DB 7; Length 435;
Best Local Similarity 20.5%; Pred. No. 1.1;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;
QY 1 MTEIVADKTVVKNVNAITAGALDLYNKYLDQVTPWQTFDETIKELSRFKQESQASV 60
Db 99 LDNLLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKTLMDSQDKYFEATQTVYEWCGVATQLLAAYILLDFEYNEKKAQAQDILIKVL 120
Db 156 L-RDVTQAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQSRLDVTQA-L 205
QY 121 DDGITKLENAQKSL-L-VSSQSFNNASGKLAL-DSQLTN-DPSEKSSYFQSDVKIRKEA 177
Db 206 ESQVEKYNDTAQSLRDVSAQLESYKSTLKEIEDLLENLTLOEKVAMAEKSVEDVQQOI 265
QY 178 YAGAAAGVAGPFGLLIISYIAAGVVEGKLIPELKNK-----LKSQVNF-----TTLS 226
Db 266 LTAESTNQ-----EYA-----RMVQDLQNRSTLKEEIKETSSFLEKITDLK 308
QY 227 NTVQKANKDI-----DAKLLKLTTEIAAIGEIKETETTRFFVVDYDDL 269
Db 309 NQLRQODEDFRKLEEGKRTAEKENVMTELT-----MEINKWRLLYEEL 353
RESULT 35
ABR56999
ID ABR56999 standard; protein; 477 AA.
XX
XX ABR56999;
XX
XX 05-AUG-2003 (first entry)
XX
XX Mouse RHAMM V-2 region amino acid sequence SEQ ID NO:77.
XX
XX RHAMM; receptor hyaluronic acid mediated motility; hyaluronic acid; HA;
XX neuroprotective; vasotropic; antidiabetic; nootropic; antiarthritic;
XX antiparkinsonian; antiinflammatory; osteopathic; dermatological; vaccine;
KW

antiarteriosclerotic; cytostatic; vulnerary; hyaluronic acid antagonist; HA binding inhibitor; multiple sclerosis; restenosis; diabetes mellitus; neurological disease; Alzheimer's disease; Parkinson's disease; cancer; arthritis; inflammatory joint disease; osteoporosis; atherosclerosis; inflammatory dermatosis; inflammatory bowel disease; metastasis; inflammatory disease; respiratory distress syndrome; wound healing; RHAMM binding protein; RABP.

Mus sp.

WO2003033535-A2.

24-APR-2003.

15-OCT-2002; 2002WO-CA001563.

15-OCT-2001; 2001US-00978309.

(TRAN-) TRANSITION THERAPEUTICS INC.

Cruz T, Pastrak A, Turley EA;

WPI; 2003-372323/35.

New P16, P-16d, human and murine P32, S3, S7, and V2 polypeptides, useful for treating e.g. multiple sclerosis, restenosis, diabetes mellitus, Alzheimer's disease, Parkinson's disease, arthritis, cancer and metastasis.

Claim 1; Page 27; 215pp; English.

The present invention describes a polypeptide (I) which binds to hyaluronic acid (HA) and thereby inhibit the binding of HA to RHAMM (receptor hyaluronic acid mediated motility), comprising an amino acid sequence of P16, P-16d, human P32, murine P32, murine S3, human S3, murine S7, human S7, murine V2 and human V2. Also described: (1) a pharmaceutical composition comprising (I); (2) an antibody which binds to (I) or to a polypeptide comprising murine V3 or human V3 amino acid sequence; (3) a vaccine composition for the treatment of multiple sclerosis, restenosis or diabetes mellitus, comprising an antigen for the antibody of (2); and (4) a method for treating multiple sclerosis, restenosis or diabetes mellitus, by administering (I), human or murine V3, or an antibody to these polypeptides. (I) has neuroprotective, vasotropic, antidiabetic, nootropic, antiparkinsonian, antiarthritic, antiinflammatory, osteopathic, dermatological, antiarteriosclerotic, cytostatic and vulnerary activities, and can be used in vaccines and as an HA antagonist. The polypeptides and the antibodies to these polypeptides are useful for treating multiple sclerosis, restenosis or diabetes mellitus, as well as neurological diseases such as Alzheimer's disease, Parkinson's disease, arthritis and other inflammatory joint diseases, osteoporosis, inflammatory dermatosis, inflammatory bowel disease and other inflammatory diseases, atherosclerosis and related diseases, cancer and metastasis, and chronic and acute respiratory distress syndrome. The polypeptides may also be used to promote wound healing. ABR56968 to ABR57022 and ACC79533 to ACC79537 represent sequences used in the exemplification of the present invention

Sequence 477 AA;

Query Match 7.4%; Score 111; DB 6; Length 477;
Best Local Similarity 20.5%; Pred. No. 1.2;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;

1 MTEIVADKTVVEVKNIAETADGNLDLYNKYLQVTPWQTFDTIKELSRFQEVSAQASV 60
99 LDNLLREKEVELEKHTAQAAILTAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
61 LVGDIKTLMDSDQKFEATQTVYECGVTATLLAAYILLPEYNEKASAKOILIKVL 120
156 L-RDVTVAQLESQEKYNDTAQSLRD---VTAQLESQ-----EKYND-TAQSLRDVTAQ-L 205
121 DDGITKLNAQSKLL-VSSQSFNNASGKLLAL-DSQLTN-DFSEKSSYFQSQVDKKEA 177

human S7, murine V2 and human V2, all derived from the protein RHAMM (Receptor hyaluronate mediated motility) a cell surface target protein that is required for activation of the AP-1 (activating protein-1) pathway on response to injury. Also included are an antibody that binds to the polypeptide cited above, a pharmaceutical or vaccine compositions for treating an inflammatory neurological disorder or diabetes mellitus (comprising the above amino acid sequence or an antigen for the antibody cited above) and treating an inflammatory bowel disease, disorder, arthritis, inflammatory lung disease, obesity, lupus, cancer, kidney fibrosis, inflammatory lung disease, obesity, lupus, cardiovascular disease, diabetes mellitus or wounds (comprising administering to a patient the above polypeptide or antibody. In the above method, the inflammatory neurological disorder is multiple sclerosis, Parkinson's disease or Alzheimer's disease. The polypeptide or antibody may also be used for diagnosing or preventing the above diseases. The present sequence is a mouse RHAMM active protein fragment.

CC human S7, murine V2 and human V2, all derived from the protein RHAMM
CC (Receptor hyaluronate mediated motility) a cell surface target
CC protein that is required for activation of the AP-1 (activating protein-
CC 1) pathway on response to injury. Also included are an antibody that
CC binds to the polypeptide cited above, a pharmaceutical or vaccine
CC compositions for treating an inflammatory neurological disorder or
CC diabetes mellitus (comprising the above amino acid sequence or an antigen
CC for the antibody cited above) and treating an inflammatory bowel disease,
CC disorder, arthritis, inflammatory lung disease, obesity, lupus,
CC cancer, kidney fibrosis, inflammatory lung disease, obesity, lupus,
CC cardiovascular disease, diabetes mellitus or wounds (comprising
CC administering to a patient the above polypeptide or antibody. In the
CC above method, the inflammatory neurological disorder is multiple
CC sclerosis, Parkinson's disease or Alzheimer's disease. The polypeptide or
CC antibody may also be used for diagnosing or preventing the above
CC diseases. The present sequence is a mouse RHAMM active protein fragment.
XX
SQ Sequence 477 AA;

Query Match 7.4%; Score 111; DB 7; Length 477;
Best Local Similarity 20.5%; Pred. No. 1.2; Indels 62; Gaps 15;
Matches 60; Conservative 63; Mismatches 63; Mismatches 63; Mismatches 63;
QY 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVVPWQTFDETIKELSRFKQYSQAASV 60
DB 99 LDNLLREKEVELEKHAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKTLMDSDQKYFEATQVYVWCGVATQLLAAYILLDFEYNEKKAQAQDILIKVL 120
DB 156 L-RDVTAQLESVQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 205
QY 121 DDGITKLNKNAQSKLL-VSSQSFNNASGKLAL-DSQLTN-DFSEKSYFQSQVDKIRKEA 177
DB 206 ESVQEKYNDTAQSLRDVTAQLESVQSKSTLKEIEDLKLENLTQEKVAMAEKSVEDVQQOI 265
QY 178 YAGAAAGVAVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
DB 266 LTAESTNQ-----EVA-----RMVQDLQNRSTLKEEIKETTSFLEKITDLK 308
QY 227 NTVKQANKDI-----DAAKLKTETIAAIGEIKETETTRFYVDYDDL 269
DB 309 NQLRQODEDFRKLQBEKGRKTAENVMTELT-----MEINKWRLLYEEL 353

RESULT 38
ABG60843
ID ABG60843 standard; protein; 631 AA.
XX
AC ABG60843;
XX
DT 13-AUG-2002 (first entry)
XX
DE Mouse receptor for hyaluronan-mediated motility (RHAMM).
XX
DE Tissue disorder; response-to-injury process; cell proliferating;
XX hyaluronate acid; HA; receptor for hyaluronan-mediated motility; RHAMM;
XX inflammatory neurological disorder; Parkinson's disease;
XX Alzheimer's disease; arthritis; multiple sclerosis; gastritis; nephritis;
XX inflammatory dermatosis; psoriasis; inflammatory bowel disease; stenosis;
XX restenosis; cancer; kidney fibrosis; inflammatory lung disease;
XX emphysema; asthma; cystic fibrosis; obesity; obesity related disease;
XX lupus; cardiovascular disease; atherosclerosis; wound; scar; diabetes;
XX tissue transplantation; stroke; inflammatory response; fibrotic response;
XX medical implant; Acquired immunodeficiency syndrome; AIDS; hepatitis;
XX myocardial fibrosis; hepatic fibrosis; chronic cystitis; acute mastitis;
XX septic shock; thyroiditis; retinopathy.
OS Mus musculus.
XX
XX WO200229415-A1.
XX
XX 11-APR-2002.
XX

Query Match 7.4%; Score 111; DB 6; Length 477;
Best Local Similarity 20.5%; Pred. No. 1.2;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;
QY 1 MTEIVADKTVVVKNAIETADGALDLYNKYLDQVVPWQTFDETIKELSRFKQYSQAASV 60
DB 99 LDNLLREKEVELEKHAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKTLMDSDQKYFEATQVYVWCGVATQLLAAYILLDFEYNEKKAQAQDILIKVL 120
DB 156 L-RDVTAQLESVQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLRDVTAQ-L 205
QY 121 DDGITKLNKNAQSKLL-VSSQSFNNASGKLAL-DSQLTN-DFSEKSYFQSQVDKIRKEA 177
DB 206 ESVQEKYNDTAQSLRDVTAQLESVQSKSTLKEIEDLKLENLTQEKVAMAEKSVEDVQQOI 265
QY 178 YAGAAAGVAVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF-----TTLS 226
DB 266 LTAESTNQ-----EVA-----RMVQDLQNRSTLKEEIKETTSFLEKITDLK 308
QY 227 NTVKQANKDI-----DAAKLKTETIAAIGEIKETETTRFYVDYDDL 269
DB 309 NQLRQODEDFRKLQBEKGRKTAENVMTELT-----MEINKWRLLYEEL 353

RESULT 37
ADC02468
ID ADC02468 standard; protein; 477 AA.
XX
AC ADC02468;
XX
DT 18-DEC-2003 (first entry)
XX
DE Mouse RHAMM V2 protein #1.
XX
DE RHAMM; Receptor hyaluronate mediated motility;
KW cell surface target protein; AP-1; activating protein-1;
KW response-to-injury; vaccine; inflammatory neurological disorder;
KW diabetes mellitus; arthritis; inflammatory dermatosis;
KW inflammatory bowel disease; cancer; kidney fibrosis;
KW inflammatory lung disease; obesity; lupus; cardiovascular disease; wound;
KW multiple sclerosis; Parkinson's disease; Alzheimer's disease;
KW alpha helical peptide; mouse.
XX
XX Mus sp.
XX
XX US2003100490-A1.
XX
XX 29-MAY-2003.
XX
XX 15-OCT-2001; 2001US-00978309.
XX
XX 01-APR-1999; 99US-0127457P.
XX
XX 03-APR-2000; 2000US-00541522.
XX
XX 05-OCT-2000; 2000US-00685010.
XX
XX (CRUZ/) CRUZ T.
XX (PAST/) PASTRAK A.
XX (TURL/) TURL E A.
XX
XX Cruz T, Pastrak A, Turley EA;
XX
XX WPI; 2003-755209/71.
XX
XX New polypeptide for diagnosing, preventing or treating inflammatory
XX neurological disorders, cancer, obesity, lupus or diabetes mellitus, S7
XX comprises an amino acid sequence of P16, P32, or murine or human S3,
XX or V2.
XX
XX Claim 1; SEQ ID NO 77; 110pp; English.
XX
XX The invention relates to a polypeptide comprising an amino acid sequence
XX selected from the sequences of P16, P32, murine S3, human S3, murine S7,
XX

SQ Sequence 794 AA;
Query Match 7.4%; Score 111; DB 5; Length 794;
Best Local Similarity 20.5%; Pred. No. 2.4;
Matches 60; Conservative 63; Mismatches 108; Indels 62; Gaps 15;
QY 1 MTEIVADKTVVKNVNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRKQVYSAASV 60
DB 416 LDNLREKEVELEKHAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 472
QY 61 LVGDIKTLMDSDQKYEATQTVVWCGVATQLLAAVILLFDEYNEKKASQAQDILIKVL 120
DB 473 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQSLRDVTAQ-L 522
QY 121 DDGITKLINEAQKSL-L-VSSQSFNNASGKLAL-DSQLTN-DFSKSSYFQSDVKIRKEA 177
DB 523 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIEDLKLNTLQEKVAMAEKSVEDVQQOI 582
QY 178 YAGAAAGVAVGPFGLIISYSIAAGVVEGKLIPELKNK-----LKSQVQNF---TTLS 226
DB 583 LTAESTNQ-----EYA-----RMVQDQNRSTLKEEIKETISFLEKITDLK 625
QY 227 NTVKQANKDI-----DAAKLTETIAAIGEIKETETTRFYVDYDDL 269
DB 626 NQLRQODEDFRQLEEKGRPTAEKENVMTLT-----MEINKWRLLYEEL 670

RESULT 40
ABU43682
ID ABU43682 standard; protein; 938 AA.
XX ABU43682;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #29209.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Staphylococcus haemolyticus.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 08-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA47552.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 71606; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway; (8)
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 938 AA;
Query Match 7.4%; Score 111; DB 6; Length 938;
Best Local Similarity 18.6%; Pred. No. 3;
Matches 60; Conservative 64; Mismatches 110; Indels 88; Gaps 11;
QY 9 TVEVKNVNAIETADGALDLYNKYLDQVTPWQTFDETIKELSRKQVYSAASVLDGDIKTL 68
DB 189 TINKIKNAVVAHDSLPQINKIADRI---EYLNHODDLDKYANGFR-----ALGNYKGD 240
QY 69 LMDSDQKYEATQTVVWCGVATQLLAAVILLFDEY-----NEKKASQAQK--- 113
DB 241 ILDAQOKLNDVNAI-----PSLNEKAKLILALNETMNPNIKLLDVASNDIPAQFPKINR 295
QY 114 --DILIKVLDDGTTKLINEAQKSLVSSQ---SPNNASGKLALDLSQTLNDFSEKSS--YF 166
DB 296 GVDIASEGFDLANTRLNDACQYLTSQAQQRVGDYQEAAGRAQEVNNOANSALRQOQSGLP 355
QY 167 QSQVDKIRKEAYAGAAAGVAVGPFGLIISYSIAAGVVEGKLIPELKNKLSKVQNFPTLS 226
DB 356 QYQIQKLSTD-----NSQDTVNDNQIVSN 379
QY 227 NTVKQANKDIDAAKLKLT-----TEIAAIGEIKETETTRFYVDYDDLMLSL-- 273
DB 380 NDVKSNNASALAEALLTSSNSDNOAKATQSDIKALKADISYGVIGSNRPTEFNDMLRNKLT 439
QY 274 -LKEAAK---KMINTCNEYOKR 291
DB 440 RLENSSEKSNQOLIDVLKELEKR 461
RESULT 41
ABB60789
ID ABB60789 standard; protein; 1013 AA.
XX ABB60789;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 9159.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX FN

CC Streptococcus pneumoniae and/or least one other GBS serotype. The
CC composition of the invention demonstrates antibacterial activity whilst
CC the polypeptides and polynucleotides may be useful in assays to diagnose
CC and identify streptococcal infections or for identifying, screening and
CC developing vaccines and other treatments for streptococcal infections.
CC The current sequence is that of a Streptococcus agalactiae ORF SAG
CC protein of the invention.
XX
SQ Sequence 278 AA;

Query Match 7.3%; Score 110; DB 8; Length 278;
Best Local Similarity 22.9%; Pred. No. 0.69;
Matches 66; Conservative 50; Mismatches 112; Indels 60; Gaps 15;

QY 41 DETIKELSRFKQ-----EYSQAASVLVGDITKLMDSDQKYPEATQTVYVCGVATOL 93
DB 2 NDAITKLSSFAEAAATLQGTAYNSAKSVATGTLTPMLOG-----MILFSETLSEKC---TEL 54

QY 94 LAAYILLFDEYN-----EKKASQKDTILIKVLDDGITKLN---EAKSLLVSSQS-FNN 143
DB 55 QTLVYVSCGDELDLSVLESKLADRSAL-KIAALLHLEHNDPEPSKSAISSTKSNIKK 113

QY 144 ASGKLALDLSQ--TNDPSEKSSYFQSQVDKIR---KEAVAGAAAGV-----VAGPGLI 193
DB 114 LKKRIKSNQKLDNLNEFNHSAATVFADISNAQSTVNQALAAVSTGFGSYNSKTFGAFKRP 173

QY 194 ISYSTAAGV-----EGKLIPELKNKLSVQNFVTLTNSVTQKANKDIDAAL-KLT 244
DB 174 TSGQWMTKTKVKNWKEREDAKAEELSKABEESKASKIENNTKNSVSDVKKLLKAA 233

QY 245 TEIAAIGELKTEETETRFVVDYDILM-----LSLKEAAK-KMINT 284
DB 234 NEAYKLGEIKKDT-----YESIISGLSNASAAALLKEVAKSKLTD 273

RESULT 47
ABU23940
ID ABU23940 standard; protein; 1163 AA.
XX
AC ABU23940;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #9467.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Clostridium acetobutylicum.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P..
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA27810.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX

PS
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway of
CC the gene product or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1163 AA;

Query Match 7.3%; Score 110; DB 6; Length 1163;
Best Local Similarity 17.9%; Pred. No. 4.9;
Matches 73; Conservative 69; Mismatches 114; Indels 152; Gaps 14;

QY 1 MTEIVADKTVVVKNAIETADG-----ALDLYNKYLDQVTPWTFDETIK----- 45
DB 109 IVDITGDE-VEVLEEGAKSVNEKCOBIIIGLSLDDFTRTV-VLPQGFSEFLKLEGKERR 165

QY 46 -----ELSRFQSVSOAASVLVGDIK-----TLLMDSQ 73
DB 166 NMLERLNLQYGDLSFLKARKIRKERENKENVLVGELKGYENINEDVLKERRELLKENN 225

QY 74 DKYFEATQTVYVWCGVATQLAAAYILLFDEYN-----EKKASQKDTILIKV-- 119
DB 226 DFFNEASK-----EYLKAEERYNEGKEVWGLQIBIEEKNRYRKOLMEKKOE 271

QY 120 --LDDGITKLENAQKSLVSSQSFNNASGKLLALDSQLTN-----DFSEKSS 164
DB 272 IDLKEKEARLGESSKVKPYDINVTNLKQIDILKEQILSRENTMKAIISLEKEDMEKKLS 331

QY 165 YFQSQVDKIRKEAVAGAAAGVAGPFGLLIISYIAGVVEGKLIPELKNLKKSVQNFPTT 224
DB 332 IAKNKEK-----ALPKFMKHIIILDAIKEKOLLDNKLEKRLQKIEK 377

QY 225 LS-----NTVKQANKDIDAOKLKTETEAIGEIKETE----- 258
DB 378 LSLSEASNKEELIKQNTKIDISLTLKIQNLESKIDNLKVPBEYKKNKINEGIFLLRNYDEKL 437

QY 259 -----TTRFVVDYD-----DLMSLLKEAAKKMINTCNFYOKR 291
DB 438 KHKNKILGDCDKFQVDFEKAASKKEMLFNKLSEERSKL-----DTYTKK 481

RESULT 48
ADE56031
ID ADE56031 standard; protein; 3187 AA.
XX

PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 17-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147200P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.

PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 7.3%; Score 109.5; DB 3; Length 366;

Best Local Similarity 22.2%; Pred. No. 1.1;
Matches 76; Conservative 41; Mismatches 130; Indels 95; Gaps 15;

Qy 7 DKTVEVVKNAIETADGALDLYNKYLDQVWPQT-----FDETIKELSRFKQBYSQ 56
Db 52 DAHMELSKHLVEVTQGVADFIETEDDV--WDNQAKYVLAYFENTKTKTLEIFKTIENC 109
Qy 57 AASVLVGDIKTLMDSDOKYFEATQTVYVWCGVATQLLAAAYILLDFBYNEKKAQAQDI- 115
Db 110 VENAEMGQL-----LITREALBF-----EKSAEKDVG 137
Qy 116 -----LIKVLDDGITKLINEAQKSLVSSQSFNNASGKLLAL-----DSQLTNDPSEK 162
Db 138 GKXKKYEKTLIED-----LKSFKEMGDPFD---GKVLTTQFELIKKQESLLEEVSET 186
Qy 163 SSVFQSOVDKIRKEAYA-----GAAAGVAGPFGLIISYSGIAAGVVEGKLIPELKNKLKS 217
Db 187 KKKIQDEITNLEKTKTLITNVVFGAFAVAVASIALIATGVGAAGFGALAAPL--LAAG 244
Qy 218 VQMFPTLSTNVQANKIDAAKUKLTETIAAIGETITE---TETTRFYVD-YDDLMSL 273
Db 245 WAGVYTTLDKKDALKQLEGLKKVIEESVEKGIKTNEEATETVSLVDGLEDRINKM 304
Qy 274 LK-----EAAKKMINTCNEYOKRHGK--KTLFEVPE 302
Db 305 LKLVDNAIDHEDNEAATRIVLT--QISKVKEKLTUKKITEVGE 344

Search completed: January 5, 2005, 10:56:33
Job time : 68.2911 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2005, 10:43:53 ; Search time 55.3174 Seconds
(without alignments)
3151.602 Million cell updates/sec

Title: US-09-993-292B-28
Perfect score: 1508
Sequence: 1 MTEIVADKTVVEVNAIETA.....TCNEYQKHCKTKLFEVPEV 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : UniProt 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1503	99.7	302	1	HLYE_ECOLI
2	1488	98.7	302	1	HLYE_ECO57
3	1395	92.5	302	1	HLYE_SALTI
4	1376	91.2	302	1	HLYE_SALPA
5	1133	75.1	300	2	Q9X288
6	563	37.3	113	1	HLEL_SHIFL
7	437	29.0	93	1	HLEL_ECOL6
8	133.5	8.9	584	1	SYR_SYNY3
9	129.5	8.6	1906	2	Q6C359
10	128	8.5	495	2	P71497
11	126	8.4	652	2	Q7CMF0
12	126	8.4	652	2	Q9X360
13	126	8.4	652	2	AAT28865
14	126	8.4	979	2	Q9EP71
15	126	8.4	992	2	Q6ZPT6
16	126	8.4	992	2	BAC98143
17	125	8.3	504	2	O8I752
18	125	8.3	2749	2	Q7REY3
19	124.5	8.3	1081	2	Q73KC5
20	124.5	8.3	1081	2	AAS12812
21	123.5	8.2	465	2	Q73AJ3
22	123.5	8.2	465	2	AAS40708
23	123.5	8.2	713	2	Q6MT03
24	123.5	8.2	713	2	CAE77235
25	123	8.2	541	2	Q74DE2
26	123	8.2	541	2	AAR34750
27	123	8.2	649	2	Q9LVQ4
28	123	8.2	1189	2	Q8CP10
29	123	8.2	2752	2	Q9BJ70
30	122.5	8.1	927	2	Q71X69
31	122.5	8.1	927	2	AAT05096

ALIGNMENTS

RESULT 1

HLYE_ECOLI	HLYE_ECOLI	STANDARD;	PRT;	302 AA.
ID	P77335; Q47276; Q8VU70; Q9R3G4;			
AC	01-NOV-1997 (Rel. 35, Created)			
DT	29-MAR-2004 (Rel. 43, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	Hemolysin E, chromosomal (Hemolysis-inducing protein) (Silent			
DE	hemolysin sheA) (Cytotoxin clyA) (Latent pore-forming 34 kDa			
DE	haemolysin).			
GN	Names: hlyE; Synonyms: clyA, sheA, hpx; OrderedLocusNames=hlyE2;			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / XL1-BLUE;			
RA	McNamara P.J., Iandolo J.J., Uhlich G.;			
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.			
RP	[2]			
RC	SEQUENCE FROM N.A.			
RA	del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;			
RT	"The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted			
RL	haemolysin.";			
RL	Mol. Microbiol. 25:107-115(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 1-12, SUBCELLULAR LOCATION, AND			
RP	INDUCTION.			
RC	STRAIN=K12;			
RX	MEDLINE=99157562; PubMed=10027972;			
RA	Ludwig A., Bauer S., Benz R., Bergmann B., Goebel W.;			
RT	"Analysis of the SlyA-controlled expression, subcellular localization			
RT	and pore-forming activity of a 34 kDa haemolysin (ClyA) from			
RT	Escherichia coli K-12.";			
RL	Mol. Microbiol. 31:557-567(1999).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=3030-2;			
RA	Xing J., Fernandez S.V., Kapur V., Barletta R.G., Moxley R.A.;			
RL	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MGI655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			

Q7shz4 neurospora
Q8i754 plasmodium
Q9szk7 arabidopsis
Q7rfq7 plasmodium
Q7raw8 plasmodium
Q8n592 homo sapien
Q60763 homo sapien
Q7rd43 plasmodium
Q8gne7 xenopus lae
P29616 gallus gall
Q700S8
Caf332691 mycoplasma
Q6cxq4 kluyveromyc
Q86tb8 homo sapien
Q8bq9 debaryomyce
Q9liw7 oryza sativ
Q7rel0 plasmodium
Q7p7e4 fusobacteri
Q895w1 clostridium

32 122.5 8.1 4007 2 Q7SHZ4
33 122 8.1 852 2 Q8I754
34 122 8.1 1496 2 Q9SZK7
35 121.5 8.1 1081 2 Q7RFQ7
36 121 8.0 1116 2 Q7RAW8
37 120.5 8.0 961 2 Q8N592
38 120.5 8.0 962 1 VDP_HUMAN
39 120.5 8.0 1611 2 Q7RD43
40 120 8.0 1012 2 Q6GNE7
41 120 8.0 1102 1 MYSC_CHICK
42 120 8.0 1828 2 Q700S8
43 120 8.0 1828 2 CAF332691
44 119.5 7.9 655 2 Q6CXQ4
45 119.5 7.9 973 2 Q86TB8
46 119.5 7.9 2042 2 Q6BQ09
47 119 7.9 895 2 Q9LIW7
48 119 7.9 2664 2 Q7RELO
49 118.5 7.9 335 2 Q7P7E4
50 118 7.8 495 2 Q895W1

RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12."; Science 277:1453-1474 (1997).
RN [6]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RN MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map."; DNA Res. 3:137-155 (1996).
RN [7]
RP SEQUENCE FROM N.A.
RX STRAIN=CH9802;
RA Chang G.-N., Ho K.-C.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RP [8]
RX SEQUENCE OF 1-295 FROM N.A.
RA STRAIN=K12 / XLI-BLUE;
RN King C.H., Shinnick T.M.;
RP Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [9]
RX SEQUENCE OF 1-155 FROM N.A.
RA STRAIN=K12 / AB1157;
RN Woodgate R.;
RP Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [10]
RX SEQUENCE OF 1-17, AND MUTAGENESIS OF 87-GLY--VAL-89; 142-ASN-ALA-143; 182-ALA-GLY-183; 186-ALA-GLY-187; ASP-267 AND 292-GLY-LYS-293.
RN MEDLINE=99316011; PubMed=10383763;
RX Oscarsson J., Mizunoe Y., Li L., Lai X.-H., Wieslander A., Uhlin B.E.;
RA "Molecular analysis of the cysC gene product of *Escherichia coli* from *Escherichia coli*."; Mol. Microbiol. 32:1226-1238 (1999).
RN [11]
RX PARTIAL SEQUENCE, FUNCTION, SUBCELLULAR LOCATION, OLIGOMERIZATION, AND DISULFIDE BOND FORMATION.
RA MEDLINE=22894281; PubMed=14532000;
RX Wai S.N., Lindmark B., Soederblom T., Takade A., Westermark M., Oscarsson J., Jass J., Richter-Dahlfors A., Mizunoe Y., Uhlin B.E.;
RA "Vesicle-mediated export and assembly of pore-forming oligomers of the enterobacterial clyA cytotoxin."; Cell 115:25-35 (2003).
RN [12]
RX MASS SPECTROMETRY, DISULFIDE BOND, AND MUTAGENESIS OF TYR-96; ASN-156; TYR-164 AND ARG-260.
RA MEDLINE=20576278; PubMed=11006277; DOI=10.1074/jbc.M005420200;
RX Atkins A., Wyborn N.R., Wallace A.J., Stillman T.J., Black L.K., Fielding A.B., Hisakado M., Artymiuk P.J., Green J.;
RA "Structure-function relationships of a novel bacterial toxin, hemolysin E. The role of alpha G."; J. Biol. Chem. 275:41150-41155 (2000).
RN [13]
RX INDUCTION.
RA MEDLINE=20507803; PubMed=11053378;
RX Westermark M., Oscarsson J., Mizunoe Y., Urbanaviciene J., Uhlin B.E.;
RA "Silencing and activation of clyA cytotoxin expression in *Escherichia coli*."; J. Bacteriol. 182:6347-6357 (2000).
RN [14]
RX INDUCTION.
RA MEDLINE=22053220; PubMed=12057949;
RA Spory A., Bosserhoff A., von Rhein C., Goebel W., Ludwig A.;
RT "Differential regulation of multiple proteins of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium by the transcriptional regulator SlyA."; J. Bacteriol. 184:3549-3559 (2002).
RN [15]
RP MUTANT PMWK16 DEL.
RX MEDLINE=22829856; PubMed=12949101;
RA Wai S.N., Westermark M., Oscarsson J., Jass J., Maier E., Benz R., Uhlin B.E.;
RT "Characterization of dominantly negative mutant clyA cytotoxin proteins in *Escherichia coli*."; J. Bacteriol. 185:5491-5499 (2003).
RN [16]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RA MEDLINE=20123445; PubMed=10560049;
RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A., Green J., Artymiuk P.J.;
RA "E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy."; Cell 100:265-276 (2000).
RN [17]
RX FUNCTION: Toxin, which has some hemolytic activity towards mammalian cells. Acts by forming a pore-like structure upon contact with mammalian cells.
CC -1- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a monomer, while in outer membrane vesicles, it oligomerizes to form a pore structure that is active. Probably forms an octamer.
CC -1- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer membrane vesicles. Also found in the periplasmic space.
CC -1- INDUCTION: During anaerobic growth. Weakly or not expressed in most strains. It is activated by slyA, while it is silenced by H-NS. Its expression is also regulated by CRP and FNR.
CC -1- PTM: In periplasm, it forms a disulfide bond between Cys-86 and Cys-284, which prevents the oligomerization. In outer membrane vesicles, the redox status prevents formation of the disulfide bond, leading to oligomerization and pore formation.
CC -1- MASS SPECTROMETRY: MW=34940; METHOD=Electrospray; RANGE=1-302; NOTE=Ref.12.
CC -1- SIMILARITY: Belongs to the hemolysin E family.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; U57430; AAB07048.1; ALT INIT.
CC EMBL; X98615; CAA67204.1; ALT_INIT.
CC EMBL; AJ001829; CAA05035.1; --
CC EMBL; U73842; AAD04731.1; --
CC EMBL; AE000216; AAC74266.1; ALT_INIT.
CC EMBL; D90751; BAA36016.1; ALT INIT.
CC EMBL; D90752; BAA36029.1; ALT INIT.
CC EMBL; D90753; BAA36037.1; ALT_INIT.
CC EMBL; AF240780; AAL55667.1; --
CC EMBL; U22466; AAA92081.1; --
CC EMBL; U13610; -- NOT_ANNOTATED_CDS.
CC PDB; 1QOY; X-ray; A=1-302.
CC EcoBASE; EB3032; --
CC EcoGene; EG13243; hlyE.
CC InterPro; IPR010356; HlyE.
CC Pfam; PF06109; HlyE; 1.
CC 3D-structure; Complete proteome; Cytolysis; Direct protein sequencing; Hemolysis; Toxin; Transmembrane.
KW INIT MET 0
FT TRANSMEM 182 202 Potential.
FT DISULFID 86 284 In monomeric form.
FT VARIANT 174 174 K -> R (in strain CH9802).
FT VARIANT 200 200 G -> A (in strain CH9802).
FT MUTAGEN 87 89 GVA->DND: Abolishes cytotoxic activity.
FT MUTAGEN 96 96 Y->H: Strongly reduces cytotoxic activity.
FT MUTAGEN 142 143 NA->DD: Abolishes cytotoxic activity.
FT MUTAGEN 156 156 N->H: Strongly reduces cytotoxic activity.
FT MUTAGEN 164 164 Y->C: Strongly reduces cytotoxic activity.

```

FT MUTAGEN 182 185 Missing: In PMWK16; retained in cytosol.
FT MUTAGEN 182 183 Loss of function.
FT MUTAGEN 182 183 AG-DD: Abolishes cytotoxic activity.
FT MUTAGEN 186 187 AG-DD: Abolishes cytotoxic activity.
FT MUTAGEN 260 260 R->K: Strongly reduces cytotoxic activity.
FT MUTAGEN 267 267 D->A: Strongly reduces cytotoxic activity.

Query Match 99.7%; Score 1503; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 5.2e-87;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 61
Db 1 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 60

Qy 62 VGDITKLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAOKDILIKVLD 121
Db 61 VGDITKLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAOKDILIKVLD 120

Qy 122 DGITKLENAOKSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 181
Db 121 DGITKLENAOKSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 180

Qy 182 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSQVNFPTTILSNVTKQANKDIDAAL 241
Db 181 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSQVNFPTTILSNVTKQANKDIDAAL 240

Qy 242 KLTTETAAIGETETETETTFYVDYDDMLSLKEAAKMINTCNEYQKRGKKTLPFVP 301
Db 241 KLTTETAAIGETETETETTFYVDYDDMLSLKEAAKMINTCNEYQKRGKKTLPFVP 300

Qy 302 EV 303
Db 301 EV 302

RESULT 2
HLYE ECO57
ID HLYE ECO57 STANDARD; PRT; 302 AA.
AC QREB3;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).
GN Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=z1944, ECs1677;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EHEC;
RX MEDLINE=20322319; PubMed=10865950; DOI=10.1016/S0923-2508(00)00143-1;
RA del Castillo F.J., Moreno P., del Castillo I.;
RT "Characterization of the genes encoding the SheA haemolysin in
  Escherichia coli O157:H7 and Shigella flexneri 2a.";
RL Res. Microbiol. 151:229-230(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
  Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
  Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
  Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 050952 / EHEC;

```

```

RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -I- FUNCTION: Toxin, which has some hemolytic activity towards
  mammalian cells. Acts by forming a pore-like structure upon
  contact with mammalian cells (By similarity).
CC -I- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a
  monomer, while in outer membrane vesicles, it oligomerizes to form
  a pore structure that is active. Probably forms an octamer (By
  similarity).
CC -I- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer
  membrane vesicles. Also found in the periplasmic space (By
  similarity).
CC -I- PTM: In periplasm, it forms a disulfide bond, which prevents the
  oligomerization. In outer membrane vesicles, the redox status
  prevents formation of the disulfide bond, leading to
  oligomerization and pore formation (By similarity).
CC -I- SIMILARITY: Belongs to the hemolysin E family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ238954; CAB64962.1; ALT_INIT.
DR EMBL; AE005335; AAG56033.1; ALT_INIT.
DR EMBL; AP002555; BAB35100.1; ALT_INIT.
DR PIR; E85696; E85696.
DR PIR; E90838; E90838.
DR HSP; P77335; IQOY.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
DR Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
FT INIT MET 0 By similarity.
FT TRANSMEM 182 202 Potential.
FT DISULFID 86 284 In monomeric form (By similarity).
SQ SSQUENCE 302 AA; 33585 MW; F261E29E1DE5FC87 CRC64;

Query Match 98.7%; Score 1488; DB 1; Length 302;
Best Local Similarity 99.0%; Pred. No. 4.6e-86;
Matches 299; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 61
Db 1 TEIVADKTVEVVKNAIETADGALDLYNKYLDQVLPWQTFDETIKELSRFQKQESQAASVL 60

Qy 62 VGDITKLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAOKDILIKVLD 121
Db 61 VGNIKTLMDSDQKYFEATQTVYEWCGVATQLLAAAYILLFDEYNEKKAQAOKDILIKVLD 120

Qy 122 DGITKLENAOKSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 181
Db 121 DGITKLENAOKSLVSSQSFNNASGKLLDLSQTLNDFSEKSSYFQSQVDKIRKEAYAGA 180

Qy 182 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSQVNFPTTILSNVTKQANKDIDAAL 241
Db 181 AAGVAGPFGGLIISYSIAAGVVEGKLIPELKNKLSQVNFPTTILSNVTKQANKDIDAAL 240

Qy 242 KLTTETAAIGETETETETTFYVDYDDMLSLKEAAKMINTCNEYQKRGKKTLPFVP 301
Db 241 KLTTETAAIGETETETETTFYVDYDDMLSLKEAAKMINTCNEYQKRGKKTLPFVP 300

Qy 302 EV 303
Db 301 EV 302

```

DR EMBL; AJ313034; CAC38363.1; --
DR EMBL; AL627270; CAD01758.1; ALT_INIT.
DR EMBL; AE016839; AAO69115.1; ALT_INIT.
DR HSP; P77335; 100Y.
DR InterPro; IPR010356; HlyE.
DR Pfam; PF06109; HlyE; 1.
KW Complete proteome; Cytolysis; Hemolysis; Toxin; Transmembrane.
FT INIT MET 0 By similarity.
FT TRANSMEM 178 198 Potential.
FT DISULFID 86 284 In monomeric form (By similarity).
SQ SEQUENCE 302 AA; 33658 MW; 2BDFD835D044FDAE CRC64;

Query Match 92.5%; Score 1395; DB 1; Length 302;
Best Local Similarity 91.1%; Pred. NO. 3.4e-80;
Matches 275; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 2 TEIVADKTVEVVKNAETADGALDLYNKYLQDVIPWQTFDETIKELSRFKQEYSQAASVL 61
DB 1 TGIFAEQTVVEVVKSAETADGALDLYNKYLQDVIPWQTFDETIKELSRFKQEYSQAASVL 60

QY 62 VGDIKTLMDSDQKYFEATQTVYVCGVATQLLAAAYILLFDEYNEKKASAKDILITVLD 121
DB 61 VGDIKVLLMDSDQKYFEATQTVYVCGVATQLLAAAYILLFDEYNEKKASAKDILITVLD 120

QY 122 DGIITKLENAQKSLIVSSQSFNNASGKLIALDSQLTNDSEKSSYFQSQVDKIRKEAYAGA 181
DB 121 DGVKKLENAQKSLITSSQSFNNASGKLIALDSQLTNDSEKSSYFQSQVDKIRKEAYAGA 180

QY 182 AAGVAGPGLIISYSIAAGVVEGKLIPELNKLSQVQNTFTLSNTVKQAKNDIDAAL 241
DB 181 AAGVAGPGLIISYSIAAGVVEGKLIPELNKLSQVQNTFTLSNTVKQAKNDIDAAL 240

QY 242 KLTETAAIGEIKETETETTRFVYVDVDDMLSLKKAQKQNTCNEYQKHGKKTLPFVP 301
DB 241 KLATEIAAIGEIKETETETTRFVYVDVDDMLSLKKAQKQNTCNEYQKHGKKTLPFVP 300

QY 302 EV 303
DB 301 DV 302

RESULT 4
HLYE SALPA
ID -HLYE SALPA STANDARD; PRT; 302 AA.
AC Q93RR6; 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).
GN Name=hlyE; Synonyms=clyA, sheA;
OS Salmonella paratyphi-a
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=54388;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMI S2068/99;
RX MEDLINE=22215712; PubMed=12228306;
RA Oscarson J., Westmark M., Loeferdahl S., Olsen B., Palmgren H.,
RA Mizunoe Y., Wai S.N., Uhlin B.E.;
RT "Characterization of a pore-forming cytotoxin expressed by Salmonella
RL Infect. Immun. 70:5759-5769(2002).
CC -!- FUNCTION: Toxin, which has some hemolytic activity towards
CC mammalian cells. Acts by forming a pore-like structure upon
CC contact with mammalian cells (By similarity).
CC -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a
CC monomer, while in outer membrane vesicles, it oligomerizes to form
CC a pore structure that is active. Probably forms an octamer (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer
CC membrane vesicles. Also found in the periplasmic space (By
CC similarity).
CC

RESULT 3
HLYE SALTI
ID -HLYE SALTI STANDARD; PRT; 302 AA.
AC Q82727; Q934C4; 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hemolysin E (Cytotoxin clyA) (Silent hemolysin sheA).
GN Name=hlyE; Synonyms=clyA, sheA; OrderedLocusNames=STY1498, t1477;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty21a and SMI S2369/96;
RX MEDLINE=22215712; PubMed=12228306;
RA Oscarson J., Westmark M., Loeferdahl S., Olsen B., Palmgren H.,
RA Mizunoe Y., Wai S.N., Uhlin B.E.;
RT "Characterization of a pore-forming cytotoxin expressed by Salmonella
RL Infect. Immun. 70:5759-5769(2002).
CC -!- FUNCTION: Toxin, which has some hemolytic activity towards
CC mammalian cells. Acts by forming a pore-like structure upon
CC contact with mammalian cells (By similarity).
CC -!- SUBUNIT: Monomer and oligomer. In periplasm, it is present as a
CC monomer, while in outer membrane vesicles, it oligomerizes to form
CC a pore structure that is active. Probably forms an octamer (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Exported from the cell by outer
CC membrane vesicles. Also found in the periplasmic space (By
CC similarity).
CC -!- PTM: In periplasm, it forms a disulfide bond, which prevents the
CC oligomerization. In outer membrane vesicles, the redox status
CC prevents formation of the disulfide bond, leading to
CC oligomerization and pore formation (By similarity).
CC -!- SIMILARITY: Belongs to the hemolysin E family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ313032; CAC38360.1; --
CC

RL Nucleic Acids Res. 30:4432-4441(2002).
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T";
 RL Infect. Immun. 71:2775-2786(2003).
 CC -1- SIMILARITY: Belongs to the hemolysin E family.
 CC from E.coli K-12 strain, it lacks all the C-terminal part of the
 CC protein, due to a deletion that creates a frameshift, and it is
 CC therefore not functional. May be a pseudogene.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF200955; AAF13995.1; -
 CC EMBL; AJ238955; CAB65415.1; ALT INIT.
 CC EMBL; AE015143; -; NOT_ANNOTATED_CDS.
 CC EMBL; AE016982; -; NOT_ANNOTATED_CDS.
 CC HSP; P77335; 100Y.
 CC InterPro; IPR010356; HlyE.
 CC Pfam; PF06109; HlyE; 1.
 CC SEQUENCE 113 AA; 12879 MW; C328908D14C54EB CRC64;
 Query Match 37.3%; Score 563; DB 1; Length 113;
 Best Local Similarity 100.0%; Pred. NO. 3.5e-28;
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTEIVADKTVVEVKNAIETADGALDLYNKYLDQVWPQDTETIKELSRFKQESQASV 60
 DB 1 MTEIVADKTVVEVKNAIETADGALDLYNKYLDQVWPQDTETIKELSRFKQESQASV 60
 QY 61 LVGDIKLLMDSDQKYFEATQTVYEWGCVATQLLAAVILLFDEYNEKKASA 111
 DB 61 LVGDIKLLMDSDQKYFEATQTVYEWGCVATQLLAAVILLFDEYNEKKASA 111
 RESULT 7
 HLEL ECOL6 STANDARD; PRT; 93 AA.
 AC Q8FI27;
 DT 29-MAR-2004 (Rel. 43, Created)
 DT 29-MAR-2004 (Rel. 43, Last sequence update)
 DE Hemolysin E-like protein.
 GN OrderedLocusNames=cl630;
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;
 RX MEDLINE=2238234; PubMed=12471157; DOI=10.1073/pnas.252529799;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -1- SIMILARITY: Belongs to the hemolysin E family.

CC -1- CAUTION: Although it is strongly related to the hemolysin E toxin
 CC from E.coli K-12 strain, it lacks all the N-terminal part of the
 CC protein, and it is therefore probably not functional. May be a
 CC pseudogene.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE016759; AAN80095.1; -
 CC HSP; P77335; 100Y.
 CC InterPro; IPR010356; HlyE.
 CC Pfam; PF06109; HlyE; 1.
 CC Complete proteome.
 CC SEQUENCE 93 AA; 10590 MW; CB1625028DC36DB6 CRC64;
 Query Match 29.0%; Score 437; DB 1; Length 93;
 Best Local Similarity 92.5%; Pred. NO. 2.5e-20;
 Matches 86; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 211 LKMKLSVQNFVFTLNTVQKQNDIAAKLKTTEIAIGETETETTRFYVDVDDL 270
 DB 1 MKMKLSALSFFTLNTVQKQNDIAAKLKTTEIAIGETETETTRFYVDVDDL 60
 QY 271 LSLLEKAAKQKQNTCNEYQKRGKTKLFEVPEV 303
 DB 61 LSLLEKAAKQKQNTCNEYQKRGKTKLFEVPEV 93
 RESULT 8
 SYR SYNY3 STANDARD; PRT; 584 AA.
 ID Q55486;
 AC Q55486;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
 GN Name=ARGS; OrderedLocusNames=sll10502;
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Suzuki T., Miyajima N.,
 RA Sugita M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 64% to 92% of the genome.";
 RL DNA Res. 2:153-166(1995).
 CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
 CC diphosphate + L-arginyl-tRNA(Arg).
 CC -1- SUBUNIT: Monomer (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D64006; BAA10833.1; -
 CC PIR; S75986; S75986.
 CC HSP; Q93RP5; 1IQO.
 CC HAMAP; MF 00123; -; 1.
 CC InterPro; IPR005148; ArgtrnAsynthet_N.

Query Match		8.5%; Score 128; DB 2; Length 495;
Best Local Similarity		23.8%; Pred. No. 5.6;
Matches		77; Conservative 47; Mismatches 119; Indels 80; Gaps 15;
QY	2 TEIVADTKVEVKNAIETADGALDLYNKYLDQVWPQTF-----DETIKELSR 49	
Db	178 TGIIVDKLISEKERLNLNLELLKKYVGQDQAIKAVTSAIMSRSGIKNPKPIGSFLF 237	
QY	50 F-----KQYSQA-ASVLVGDITKLLMSDQDKYFEATQV-----YEWCGVAT 91	
Db	238 FGPTGVGKTEVARSLADILFNSPKMIRLDMSEYMERKSHVAKLIGAPPVGVGEGRLT 297	
QY	92 QLLAA---YILLDEYNEKASAKDKILIKVLDDG-----ITKLNBAOKSLLVSSQSFFN 143	
Db	298 EAVARNPYSILFDEI-EKAHSDVFNILLQLDDGRLTDSLGKTIIDFKNTIIWTS--NI 354	
QY	144 ASGKLLALDSQLTNDFSE-KSSYFQSQVDKIRKEAYAGAAAGVAGVPGFLIISYSAAGV 202	
Db	355 ASQYLLTSD-----DFVQIDDKIQAEINQTFRPEFLNRIDNIVY--FNALSVQTI--GE 405	
QY	203 VEGKLIPELKNKLSQVNFPTLNTVQKQANDIDAAKLTTEIAAIGETETTRF 262	
Db	406 IVDKLLDELITRLQDEQNYFNFS-----ERARNKIINE----- 439	
QY	263 YVDYDDLML-LSLLKEAAKMINT 284	
Db	440 --GYDRLFGARLLKDLKLIET 460	
RESULT 11		
Q7CMF0 PRELIMINARY; PRT; 652 AA.		
AC	Q7CMF0	
DT	05-JUL-2004 (TrEMBLrel. 27, Created)	
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)	
DE	S-layer protein, (PX01-90).	
DE	Name=BXA0124;	
GN	Bacillus anthracis str. A2012.	
OG	Plasmid pX01.	
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
OX	NCBI_TaxID=191218;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A2012;	
RX	MEDLINE=22061436; PubMed=12004073;	
RA	Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,	
RA	Holtzapple E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,	
RA	Keim P., Fraser C.M.;	
RT	"Comparative Genome Sequencing for Discovery of Novel Polymorphisms in	
RT	Bacillus anthracis."	
RL	Science 296:2028-2033(2002).	
DR	EMBL; AE011190; AM26077.1; --	
DR	InterPro; IPR001119; SLH.	
DR	InterPro; IPR010989; t-snare.	
DR	InterPro; IPR010978; tRNA_binding_arm.	
DR	Pfam; PF00395; SLH; 3.	
DR	PROSITE; PS01072; SLH_DOMAIN; UNKNOWN_1.	
KW	Plasmid.	
SQ	SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;	
Query Match		8.4%; Score 126; DB 2; Length 652;
Best Local Similarity		20.5%; Pred. No. 10;
Matches		68; Conservative 57; Mismatches 108; Indels 98; Gaps 11;
QY	7 DKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQYSQAASVLVGDIK 66	
Db	233 NKRRDVINQKIDFED-KLNSQRDLERML-----EELNQLSKLQKQSPQ-----LQDLK 281	
QY	67 TLLMSDQDKYFEATQ-----TVYEWCGVATQLLAAAYILLDFEYNE--KKASAKD 114	
Db	282 NKLKESQSRLLLELNKDSNRLEINSEIKLNDRAKELLSLIMELIKQOSEPDKIKNEKD 341	
QY	115 ILIKVLDGDTITKLNBAOK-----SLVSSQSFFNNAAGKLLALDSQLTNDFSE 161	
Db	342 DLNKKREDLINRIAESKELAKKAEALNTKLVLFVKQAEALNKKSQGLYYINKLDNELRE 401	
RESULT 12		
Q9X360 PRELIMINARY; PRT; 652 AA.		
ID	Q9X360	
AC	Q9X360;	
DT	01-NOV-1999 (TrEMBLrel. 12, Created)	
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)	
DT	01-OCT-2004 (TrEMBLrel. 28, Last annotation update)	
DE	PX01-90 (S-layer protein,).	
GN	OrderedLocusNames=GBAA_PX01_0124;	
OS	Bacillus anthracis.	
OG	Plasmid virulence plasmid pX01, and Plasmid pX01.	
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
OX	NCBI_TaxID=1392;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Stjerne; PLASMID=virulence plasmid pX01;	
RX	MEDLINE=99445483; PubMed=10515943;	
RA	Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,	
RA	Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,	
RA	Martinez Y., Ricke D., Svensson R., Jackson P.J.;	
RT	"Sequence and organization of pX01, the large Bacillus anthracis	
RT	plasmid harboring the anthrax toxin genes."	
RL	J. Bacteriol. 181:6509-6515(1999).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Ames / isolate 0581; PLASMID=pX01;	
RA	Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,	
RA	Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,	
RA	Fraser C.M.;	
RT	"Bacillus anthracis comparative genomics."	
RT	Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AF065404; AAD32394.1; --	
DR	EMBL; AE017336; AAT28865.2; --	
DR	PIR; B59102; B59102.	
DR	TIGR; GBAA_PX01_0124; --	
DR	InterPro; IPR001119; SLH.	
DR	Pfam; PF00395; SLH; 3.	
DR	PROSITE; PS01072; SLH_DOMAIN; UNKNOWN_1.	
KW	Complete proteome; Plasmid.	
SQ	SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;	
Query Match		8.4%; Score 126; DB 2; Length 652;
Best Local Similarity		20.5%; Pred. No. 10;
Matches		68; Conservative 57; Mismatches 108; Indels 98; Gaps 11;
QY	7 DKTVEVVKNAIETADGALDLYNKYLDQVWPQTFDETIKELSRFKQYSQAASVLVGDIK 66	
Db	233 NKRRDVINQKIDFED-KLNSQRDLERML-----EELNQLSKLQKQSPQ-----LQDLK 281	
QY	67 TLLMSDQDKYFEATQ-----TVYEWCGVATQLLAAAYILLDFEYNE--KKASAKD 114	
Db	282 NKLKESQSRLLLELNKDSNRLEINSEIKLNDRAKELLSLIMELIKQOSEPDKIKNEKD 341	
QY	115 ILIKVLDGDTITKLNBAOK-----SLVSSQSFFNNAAGKLLALDSQLTNDFSE 161	
Db	342 DLNKKREDLINRIAESKELAKKAEALNTKLVLFVKQAEALNKKSQGLYYINKLDNELRE 401	

```
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSTAAGVVEGKLIPELKNKLKSVQNF 221
D 402 LADKYNSDNKISR-----LKNHIGYENKQ 426
QY 222 FTTLSNTVKQANKIDAAKLLKLTETIAAIGBIKTETTRFYVDYDDMLSLKKEAAKQM 281
D 427 LEKIELEECNKKIDNTKKQL-AEPDKSNKKQOELE-----SELVOLNKKI 472
QY 282 INTCNVEYQKRH-----GKTKLFVPEV 303
D 473 ----DELGKGRKHQOELEASQKKALDEAKEI 499
RESULT 13
ID AAT28865 PRELIMINARY; PRT; 652 AA.
AC AAT28865;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DE S-layer protein, (PXOI-90).
GN GBAA_PXOI_0124.
OS Bacillus anthracis str. Ames 0581.
OG Plasmid pXOI.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581.
RA Ravel J., Rasko D.A., Shumway M.P., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE017336; AAT28865.2; -.
KW Plasmid
SQ SEQUENCE 652 AA; 76210 MW; 723F5FB03516355 CRC64;
Query Match 8.4%; Score 126; DB 2; Length 652;
Best Local Similarity 20.5%; Pred. No. 10;
Matches 68; Conservative 57; Mismatches 108; Indels 98; Gaps 11;
QY 7 DKTVEVVKNAIETADGALDLYNKYLDQVIFQWTFDETIKELSRPKQYSAASVLVDGIK 66
D 233 NKERDVINQKIDBFD-KLNSQRKLERML-----BELNQKLSQLKQSQPQ-----LQDLK 281
QY 67 TLLMDSQDKYPEATO-----TVYEWGVATQLLAAYVILLFDYNE--KKASAQKD 114
D 282 NKLESQSRLELKNKDSNLELNSEIKLNDKRAELSLIMELIKQSEFDKIRKNERD 341
QY 115 ILIKVLDDGITKLNEAQK-----SLIVSSQSFNNASGKLLALDLSQLTNDPSE 161
D 342 DLNKKREDLNRIAESKELAKKKAELNTKLVELFKVQEAALNKKSGQVLYYINKLDNELRE 401
QY 162 KSSYFQSQVDKIRKEAYAGAAAGVAGPFGLLIISYSTAAGVVEGKLIPELKNKLKSVQNF 221
D 402 LADKYNSDNKISR-----LKNHIGYENKQ 426
QY 222 FTTLSNTVKQANKIDAAKLLKLTETIAAIGBIKTETTRFYVDYDDMLSLKKEAAKQM 281
D 427 LEKIELEECNKKIDNTKKQL-AEPDKSNKKQOELE-----SELVOLNKKI 472
QY 282 INTCNVEYQKRH-----GKTKLFVPEV 303
D 473 ----DELGKGRKHQOELEASQKKALDEAKEI 499
RESULT 14
Q9EP71
ID Q9EP71 PRELIMINARY; PRT; 979 AA.
AC Q9EP71;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
```

```
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE NORPEG-like protein (Ankyrcorbin).
GN Name=Rail4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Kuty R.K., Kuty G., Samuel W., Duncan T., Bridges C.C.,
RA El-Sherbeeny A., Naginei C.N., Smith S.B., Wiggert B.;
RT "Molecular characterization and developmental expression of NORPEG, a
RT novel gene induced by retinoic acid.";
RL J. Biol. Chem. 276:2831-2840(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapletton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57 Black/6;
RA MEDLINE=21110708; PubMed=11168586;
RA Peng Y.F., Mandai K., Sakisaka T., Okabe N., Yamamoto Y., Yokoyama S.,
RA Mizoguchi A., Shiozaki H., Monden M., Takai Y.;
RT "Ankyrcorbin: a novel actin cytoskeleton-associated protein.";
RL Genes Cells 5:1001-1008(2000).
DR EMBL: AF274866; AAG25937.1; -.
DR EMBL: BC052458; AAG25458.1; -.
DR EMBL: AF202315; AAG24483.1; -.
DR HSSP: P42773; 1IHB.
DR MGD: MGI:1922896; Rail4.
DR GO: GO:0005739; C:mitochondrion; IDA.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; Ank; 7.
DR PRINTS: PR01415; ANKYRIN.
DR SMART: SM00248; ANK; 6.
DR PROSITE: PS50088; ANK_REPEAT; 5.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat.
SQ SEQUENCE 979 AA; 108851 MW; B2E8C016D80237C4 CRC64;
Query Match 8.4%; Score 126; DB 2; Length 979;
Best Local Similarity 21.5%; Pred. No. 17;
Matches 78; Conservative 64; Mismatches 125; Indels 96; Gaps 17;
```

QY 1 MTEIVA-DKTVEVVKNAIETADGALDLYNKYLDQVIPW--QTDETIKELSRPKQYSA 57

Db	566	---	VRLETKLAKEQA	---	EATKPPAEAC	---	EELRSSYCSVIENNKKAFLEKYQQ	616
Qy	111	AQDILIKVLDDG	ITKLNEAQSLLV	---	SSQSFNNASGKLALDSQLTNDPSEKSSY	165		
Db	617	AQBEIM	---	KLKDTLSQMPQEA	PPDSDGDMKEAMNRM	---	DELNKQVSELSQL	664
Qy	166	F	---	QSQVDKIRK	EAYAGAAAGV	PPFGLI	ISYSIAAGVVEGKLIPELKNKLSQVON	220
Db	665	YREAQAELEDY	RKRKSLDAAEY	THKAHERLMHVSNL	SRAKSEALSEMKSQYSKVLNE	724		
Qy	221	---	FFTLTSLNTVKQ	---	ANKDIDAACL	---	KL	243
Db	725	LTQLKQLVDAH	KENSVSITEHLQVIT	TLRTTAKEMEEKISALT	CHLANKEAEVAKLEKQL	784		
Qy	244	TTEIAAIGEIK	TETTRFYVDYD	DLMLSLKEA	---	AKQMINTCNEYQKHGKTLFEV	300	
Db	785	AEEKAASVDM	VPKSS	---	YEKLQASLESEVNALATKLKESVREKRAHS	---	EV	833
Qy	301	PEV	303					
Db	834	AOV	836					
RESULT 16								
Db	BAC98143	PRELIMINARY;	PRT;	992	AA.			
Qy	AC	BAC98143						
Db	DT	02-MAR-2004	(TREMELrel. 27, Created)					
Qy	DT	02-MAR-2004	(TREMELrel. 27, Last sequence update)					
Db	DT	02-MAR-2004	(TREMELrel. 27, Last annotation update)					
Qy	DE	MKIAA1334	protein (Fragment).					
Db	GN	MKIAA1334						
Qy	OS	Mus musculus (Mouse).						
Db	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Qy	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
Db	OX	NCBI_TaxID=10090;						
Qy	RN	[1]						
Db	RP	SEQUENCE FROM N.A.						
Qy	RC	TISSUE=Embryonic tail;						
Db	RA	Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,						
Qy	RA	Saga Y., Nagase T., Ohara O., Koga H.; Mouse Homologues of KIAA Gene:						
Db	RT	"Prediction of the Coding Sequences of 500 Mouse KIAA-homologous						
Qy	RT	III. The Complete Nucleotide Sequences of 500 Mouse KIAA-homologous						
Db	RT	cDNAs Identified by Screening of Terminal Sequences of cDNA Clones						
Qy	RT	Randomly Sampled from Size-fractionated Libraries."						
Db	RL	DNA Res. 10:167-180(2003).						
Qy	DR	EMBL; AKI29333; BAC98143.1;						
Db	FT	NON TER						
Qy	SQ	SEQUENCE	992 AA; 110311 MW; FBBC2380E775EB5B	CRC64;				
Query Match								
Best Local Similarity	8.4%;	Score 126;	DB 2;	Length 992;				
Matches	78;	Conservative	21.5%;	Pred. No. 17;				
Qy	1	MTEIVA-DKTVVVVNAIETADGALDLYNKYLQVIPW	---	OTDETIKELSRFKQYSQA	57			
Db	510	MKEVLSVQKMKLLGSQESADGYSHLREAPADEDIDTLKQDKQAVESARKER	---	---	565			
Qy	58	ASVLVGDIKTLMDSDQKYFEATQTVYEWGCVATQLAAYILLFDEYNKKA	---	---	S 110			
Db	566	---	VRLETKLAKEQA	---	EATKPPAEAC	---	EELRSSYCSVIENNKKAFLEKYQQ	616
Qy	111	AQDILIKVLDDG	ITKLNEAQSLLV	---	SSQSFNNASGKLALDSQLTNDPSEKSSY	165		
Db	617	AQBEIM	---	KLKDTLSQMPQEA	PPDSDGDMKEAMNRM	---	DELNKQVSELSQL	664
Qy	166	F	---	QSQVDKIRK	EAYAGAAAGV	PPFGLI	ISYSIAAGVVEGKLIPELKNKLSQVON	220
Db	665	YREAQAELEDY	RKRKSLDAAEY	THKAHERLMHVSNL	SRAKSEALSEMKSQYSKVLNE	724		
Qy	221	---	FFTLTSLNTVKQ	---	ANKDIDAACL	---	KL	243

Db	497	MKEVLSVQKMKLLGSQESADGYSHLREAPADEDIDTLKQDKQAVESARKER	---	---	---	---	---	552
Qy	58	ASVLVGDIKTLMDSDQKYFEATQTVYEWGCVATQLAAYILLFDEYNKKA	---	---	---	---	---	S 110
Db	553	---	VRLETKLAKEQA	---	EATKPPAEAC	---	EELRSSYCSVIENNKKAFLEKYQQ	603
Qy	111	AQDILIKVLDDG	ITKLNEAQSLLV	---	SSQSFNNASGKLALDSQLTNDPSEKSSY	165		
Db	604	AQBEIM	---	KLKDTLSQMPQEA	PPDSDGDMKEAMNRM	---	DELNKQVSELSQL	651
Qy	166	F	---	QSQVDKIRK	EAYAGAAAGV	PPFGLI	ISYSIAAGVVEGKLIPELKNKLSQVON	220
Db	652	YREAQAELEDY	RKRKSLDAAEY	THKAHERLMHVSNL	SRAKSEALSEMKSQYSKVLNE	711		
Qy	221	---	FFTLTSLNTVKQ	---	ANKDIDAACL	---	KL	243
Db	712	LTQLKQLVDAH	KENSVSITEHLQVIT	TLRTTAKEMEEKISALT	CHLANKEAEVAKLEKQL	771		
Qy	244	TTEIAAIGEIK	TETTRFYVDYD	DLMLSLKEA	---	AKQMINTCNEYQKHGKTLFEV	300	
Db	772	AEEKAASVDM	VPKSS	---	YEKLQASLESEVNALATKLKESVREKRAHS	---	EV	820
Qy	301	PEV	303					
Db	821	AOV	823					
RESULT 15								
Q6ZPT6	PRELIMINARY;	PRT;	992	AA.				
Q6ZPT6								
DT	05-JUL-2004	(TREMELrel. 27, Created)						
DT	05-JUL-2004	(TREMELrel. 27, Last sequence update)						
DT	05-JUL-2004	(TREMELrel. 27, Last annotation update)						
DE	MKIAA1334	protein (Fragment).						
GN	Name=MKIAA1334;							
OS	Mus musculus (Mouse).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
OX	NCBI_TaxID=10090;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	TISSUE=Embryonic tail;							
RX	PubMed=14621295;							
RA	Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,							
RA	Saga Y., Nagase T., Ohara O., Koga H.;							
RT	"Prediction of the coding sequences of mouse homologues of KIAA gene:							
RT	III. the complete nucleotide sequences of 500 mouse KIAA-homologous							
RT	cDNAs identified by screening of terminal sequences of cDNA clones							
RT	randomly sampled from size-fractionated libraries."							
RL	DNA Res. 10:167-180(2003).							
DR	EMBL; AKI29333; BAC98143.1;							
DR	GO; GO:0005739; C-mitochondrion; IDA.							
DR	InterPro; IPR002110; ANK.							
DR	Pfam; PF00023; Ank; 7.							
DR	PRINTS; PR01415; ANKYRIN.							
DR	SMART; SMO0248; ANK; 6.							
DR	PROSITE; PS50088; ANK_REPEAT; 5.							
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.							
KW	ANK repeat.							
NON TER								
SQ	SEQUENCE	992 AA; 110311 MW; FBBC2380E775EB5B	CRC64;					
Query Match								
Best Local Similarity	8.4%;	Score 126;	DB 2;	Length 992;				
Matches	78;	Conservative	21.5%;	Pred. No. 17;				
Qy	1	MTEIVA-DKTVVVVNAIETADGALDLYNKYLQVIPW	---	OTDETIKELSRFKQYSQA	57			
Db	510	MKEVLSVQKMKLLGSQESADGYSHLREAPADEDIDTLKQDKQAVESARKER	---	---	565			
Qy	58	ASVLVGDIKTLMDSDQKYFEATQTVYEWGCVATQLAAYILLFDEYNKKA	---	---	S 110			

```
Db 725 LTQLKQLVDAHKENSVSITLHQLVITTTTAKEMEEKISALTCHLANKEAEVAKLEKQL 784
Qy 244 TTEIAAIGEIKTETTRFVVDLMLSLKA---AKQINTCNEYQKRGHKKTLFEV 300
Db 785 ABEKAASVDAMPKSS-----YEKLOASLESEVNALATKLKESVREKRAHS-----EV 833
Qy 301 PEV 303
Db 834 AQV 836

RESULT 17
ID Q81752 PRELIMINARY; PRT; 504 AA.
AC Q81752;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Rhoptry protein (Fragment).
GN Name-rhop235-15;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S.M., Jarra W., Bayele H., Preiser P.R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Green J.L., Holder A.A.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Bayele H.K.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ489165; CAD32955.1; -.
FT NON_TER 1
FT NON_TER 504
SQ SEQUENCE 504 AA; 58810 MW; 514CC9110674F83A CRC64;

Query Match 8.3%; Score 125; DB 2; Length 504;
Best Local Similarity 21.5%; Pred. No. 8.8;
Matches 72; Conservative 60; Mismatches 109; Indels 94; Gaps 16;

Qy 3 EIVAD-----KTVE-VVKNAIETADGALDIY---NKYLDQVPMQTFDETIKELSRF 50
Db 1 EKVDINTYNEDPKEIEKKIENVVKKIDKKNIYKEINKLNISEIEKDKTSLEELKNI 60
Qy 51 KQESQAASVLVDGIKTLMDSDQKYPEATQTVYEWCGVATQLLAAYILLFDEYNEKKAS 110
Db 61 NLSYGRS---LGNIFLEQIDEKKKAERT-----IKAMEAYIEDLDNIKKKSD 106
Qy 111 AQKDILIKVLDDGITKLNIAQKSLVSSQSFNNASGK-----LLALDSOLTNDFS 160
Db 107 IEKDMKIKM-----DINEEMKALNISDDDRNHTKSKDHKKGISDIHDKSKKIQNFS 160
Qy 161 EKS-----SYFQSQVDKIRK-----EAYAGAAGVVAGPGLIISYIAAGVVEGKLIP 209
Db 161 KESDINNKNELQENVSESRKNSDINHYSKVENI-----YNI-----1397
Qy 210 ELK-NKLKSVQNPFTLSTNTVQKANDIDAAL---KLTTETI---AAIGIKTETETT-- 260
Db 200 -LKLNKIKKIIDKVEYTDIEIKNNKNSINDELINSNGKIITKIKENSSLTCQSKIESTID 258
Qy 261 -----RFYVDYDDLMLSLKKEAAKKMINTCNEYOK 290
Db 259 DNYISKICKIADLTKYILSEE-----NNINTYLYK 288

RESULT 18
Q7REV3
ID Q7REV3 PRELIMINARY; PRT; 2749 AA.
AC Q7REV3;
```

```
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 235 kDa rhoptry protein (Fragment).
GN Name-PY04930;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., M.J.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RL parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001532; EAA16872.1; -.
DR InterPro; IPR003900; KID repeat.
DR InterPro; IPR006499; ReticulocyteBP.
DR Pfam; PF02524; KID; 6.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 2749 AA; 322590 MW; PFE10079886F63CF CRC64;

Query Match 8.3%; Score 125; DB 2; Length 2749;
Best Local Similarity 21.5%; Pred. No. 64;
Matches 72; Conservative 60; Mismatches 109; Indels 94; Gaps 16;

Qy 3 EIVAD-----KTVE-VVKNAIETADGALDIY---NKYLDQVPMQTFDETIKELSRF 50
Db 1139 EKVDINTYNEDPKEIEKKIENVVKKIDKKNIYKEINKLNISEIEKDKTSLEELKNI 1198
Qy 51 KQESQAASVLVDGIKTLMDSDQKYPEATQTVYEWCGVATQLLAAYILLFDEYNEKKAS 110
Db 1199 NLSYGRS---LGNIFLEQIDEKKKAERT-----IKAMEAYIEDLDNIKKKSD 1244
Qy 111 AQKDILIKVLDDGITKLNIAQKSLVSSQSFNNASGK-----LLALDSOLTNDFS 160
Db 1245 IEKDMKIKM-----DINEEMKALNISDDDRNHTKSKDHKKGISDIHDKSKKIQNFS 1298
Qy 161 EKS-----SYFQSQVDKIRK-----EAYAGAAGVVAGPGLIISYIAAGVVEGKLIP 209
Db 1299 KESDINNKNELQENVSESRKNSDINHYSKVENI-----YNI-----1337
Qy 210 ELK-NKLKSVQNPFTLSTNTVQKANDIDAAL---KLTTETI---AAIGIKTETETT-- 260
Db 1338 -LKLNKIKKIIDKVEYTDIEIKNNKNSINDELINSNGKIITKIKENSSLTCQSKIESTID 1396
Qy 261 -----RFYVDYDDLMLSLKKEAAKKMINTCNEYOK 290
Db 1397 DNYISKICKIADLTKYILSEE-----NNINTYLYK 1426

RESULT 19
Q73KC5
ID Q73KC5 PRELIMINARY; PRT; 1081 AA.
AC Q73KC5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
```

OrderedLocusNames=TDE2293;
 OS Treponema denticola.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 NCBI_TaxID=158;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35405 / DSM 14222;
 RX PubMed=15064399; DOI=10.1073/pnas.0307639101;
 RA Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,
 RA Dodson R.J., Davidson T.M., DeBoy R.T., Fouts D.E., Haft D.H.,
 RA Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J.F.,
 RA Durkin S.A., Daugherty S.C., Shetty J., Shvartsbeyn A.,
 RA Gebregeorgis E., Geer K., Tsagaye G., Malek J.A., Ayodeji B.,
 RA Shatsman S., McLeod M.P., Smajs D., Howell J.K., Pal S., Amin A.,
 RA Weinstein G.M., Norris S.J., Fraser C.M., Paulsen I.T.,
 RA "Comparison of the genome of the oral pathogen Treponema denticola
 RT with other spirochete genomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).
 DR EMBL; AB017253; AAS12812.1; --
 DR TIGR; TDE2293; --
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1081 AA; 123105 MW; 6490D3EA8068C18A CRC64;
 TIGR; TDE2293; --
 DR InterPro; IPR009074; Apolipo_A_E_C3.
 DR InterPro; IPR003191; GBP.
 DR InterPro; IPR009057; Homeodomain like.
 DR InterPro; IPR010992; IHF like DNA_bnd.
 DR InterPro; IPR002017; Spectrin.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1081 AA; 123105 MW; 6490D3EA8068C18A CRC64;
 Query Match 8.3%; Score 124.5; DB 2; Length 1081;
 Best Local Similarity 24.6%; Pred. No. 23;
 Matches 82; Conservative 46; Mismatches 133; Indels 73; Gaps 16;
 QY 8 KTVVVVKNNA-IETADGALDY-----NKYLDQVWPOTFDETIKELSRFKQYSSQAASVLV 62
 DB 538 KDVVAGLSRVSAMDGTVEQYIESFRADVDQ-IAKASQQLTELASYKEQVREAVSNQN 596
 QY 63 GDKITLLMDSQDKY-----FEATQTVYE-WCGVATQLLAAYILLDFEYNEKKASA 111
 DB 597 AELENTSGMGEKLLSIRESEAKFETIKKDFETWKNRTDQOFTDARSFDEKITNFAGL 656
 QY 112 QKDILIKVLDDGITKLINEAQSLLV--SSQSFNNASGKLALD-----SOLT 156
 DB 657 TEN-AIKNLD---AKYNAQYKDFVAKSGDAPNGIQAKLSVDTKVASANKAIDEHASEVT 712
 QY 157 NDFSEKSSYFQSOVDKIRKEAYAGVAGVAGPFGLLIISYSIAAGVVEGK----- 206
 DB 713 NRNTEAEKLENEVINKRIKEA-ASEAALSVOGINDMILEVRNRLDETQEKVREKIEADAD 771
 QY 207 ----LIPELKNKLKSVQNFFTLSNTVKQANKDIDAAKLKLTTETAAIGETETETTRF 262
 DB 772 RLNSLIEIDKK-----QNDFITQTKVFERA----DELKEGLEKDIASL-----KNEVTKF 818
 QY 263 YV---DYDDLMLSL-----LLKEAAKMKMINTCNE 287
 DB 819 EVYKNAMDDLTLQYDKVTHLEBEAKQKVSFRFME 852
 RESULT 20
 ID AAS12812 PRELIMINARY; PRT; 1081 AA.
 AC AAS12812;
 DT 02-WAR-2004 (TrEMBLrel. 27, Created)
 DT 02-WAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 GN TDE2293.
 OS Treponema denticola.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 NCBI_TaxID=158;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35405 / DSM 14222;

PubMed=15064399;
 RA Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,
 RA Dodson R.J., Davidson T.M., DeBoy R.T., Fouts D.E., Haft D.H.,
 RA Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J.F., Durkin S.A.,
 RA Daugherty S.C., Shetty J., Shvartsbeyn A., Gebregeorgis E., Geer K.,
 RA Tsagaye G., Malek J.A., Ayodeji B., Shatsman S., McLeod M.P.,
 RA Smajs D., Howell J.K., Pal S., Amin A., Vashisth P., McNeill T.Z.,
 RA Xiang Q., Sodergren E., Baca E., Weinstein G.M., Norris S.J.,
 RA Fraser C.M., Paulsen I.T.,
 RA "Comparison of the genome of the oral pathogen Treponema denticola
 RT with other spirochete genomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).
 DR EMBL; AB017253; AAS12812.1; --
 DR TIGR; TDE2293; --
 KW Hypothetical protein.
 SQ SEQUENCE 1081 AA; 123105 MW; 6490D3EA8068C18A CRC64;
 Query Match 8.3%; Score 124.5; DB 2; Length 1081;
 Best Local Similarity 24.6%; Pred. No. 23;
 Matches 82; Conservative 46; Mismatches 133; Indels 73; Gaps 16;
 QY 8 KTVVVVKNNA-IETADGALDY-----NKYLDQVWPOTFDETIKELSRFKQYSSQAASVLV 62
 DB 538 KDVVAGLSRVSAMDGTVEQYIESFRADVDQ-IAKASQQLTELASYKEQVREAVSNQN 596
 QY 63 GDKITLLMDSQDKY-----FEATQTVYE-WCGVATQLLAAYILLDFEYNEKKASA 111
 DB 597 AELENTSGMGEKLLSIRESEAKFETIKKDFETWKNRTDQOFTDARSFDEKITNFAGL 656
 QY 112 QKDILIKVLDDGITKLINEAQSLLV--SSQSFNNASGKLALD-----SOLT 156
 DB 657 TEN-AIKNLD---AKYNAQYKDFVAKSGDAPNGIQAKLSVDTKVASANKAIDEHASEVT 712
 QY 157 NDFSEKSSYFQSOVDKIRKEAYAGVAGVAGPFGLLIISYSIAAGVVEGK----- 206
 DB 713 NRNTEAEKLENEVINKRIKEA-ASEAALSVOGINDMILEVRNRLDETQEKVREKIEADAD 771
 QY 207 ----LIPELKNKLKSVQNFFTLSNTVKQANKDIDAAKLKLTTETAAIGETETETTRF 262
 DB 772 RLNSLIEIDKK-----QNDFITQTKVFERA----DELKEGLEKDIASL-----KNEVTKF 818
 QY 263 YV---DYDDLMLSL-----LLKEAAKMKMINTCNE 287
 DB 819 EVYKNAMDDLTLQYDKVTHLEBEAKQKVSFRFME 852
 RESULT 21
 ID Q73AJ3 PRELIMINARY; PRT; 465 AA.
 AC Q73AJ3;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Flagellin.
 GN OrderedLocusNames=BCE1780;
 OS Bacillus cereus (strain ATCC 10987).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=222523;
 [1]
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RX PubMed=14960714;
 RA Rasko D.A., Ravel J., Oekstad O.A., Helgason B., Cer R.Z., Jiang L.,
 RA Shores K.A., Fouts D.E., Tourasse N.J., Angioli S.V., Kolonay J.F.,
 RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.,
 RA "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
 RT adaptations and a large plasmid related to Bacillus anthracis pX01.";
 RL Nucleic Acids Res. 32:977-988(2004).
 DR EMBL; AE017269; AAS40708.1; --
 DR TIGR; BCE1780; --
 DR InterPro; IPR001029; Flagellin_C.
 DR InterPro; IPR001492; Flagellin_N.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR Pfam; PF00669; Flagellin_N; 1.

Qy	63	GDIKTLMSDQKVF	ATQVTVWCGVATOLLAAAYILLFDEYNK-----KA	109			
Db	180	LETKTVALSKEBAKVKV	HSB-----KIELLRKEIAAVNESVBQTKLACSQARKBOSEI	235			
Qy	110	SAQKDILIKVLDG	ITKLNIAQKSLVSSQSFNNASGKLALDSQLTNDSEKSSYFOSQ	169			
Db	236	FAKETQOKSYKAG--	MESAKSKLALKNFDPFAK--KLEVLQTLTYNE-----	282			
Qy	170	VDKIRKAYAGAAGV	VGFGIILISYIAAGVVGKLIPELKNKLSVQNFFTTLSNTV	229			
Db	283	IDELQKMET-AKAS	DISVNGVSLNEAKGLFE-KLVEEKSQBELVESLKAELKN-V	339			
Qy	230	KQANKIDDAKLLT	ETIA---AIGEIKTE-----TETTFYVDYDGLMLSLK-----	275			
Db	340	KMHDEVEAKEBIES	VAGDLHLKLSRKSSELCVTEESKAKAALDEMLLTINQISSET	399			
Qy	276	EAAK-----	MINTCNEYOK	290			
Db	400	EAARREAGNRKAK	ELMK	418			
RESULT 28							
Q8CP10							
ID	Q8CP10	PRELIMINARY;					
AC	Q8CP10	PRT; 1189 AA.					
DT	01-MAR-2003	(TEMBLrel. 23, Created)					
DT	01-MAR-2003	(TEMBLrel. 23, Last sequence update)					
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)					
DE	Chromosome segregation SMC protein.						
GN	OrderedLocusNames=SE0509;						
OS	Staphylococcus epidermidis						
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.						
OX	NCBI_TaxID=1282;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=ATCC 12228;						
RX	PubMed=12950922;						
RA	Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,						
RA	Qin Z.-Q., Zhao Y.-G., Wang W.-Y., Chen R.-S., Shen Y.,						
RA	Yuan Z.-H., Miao G.-P., Qu D., Danchin A., Wen Y.-M.;						
RT	"Genome-based analysis of virulence genes in a non-biofilm-forming						
RT	Staphylococcus epidermidis strain (ATCC 12228).";						
RL	Mol. Microbiol. 49:1577-1593(2003).						
DR	ENBL; AE016747; AA004506.1; -.						
DR	HSSP; P58301; Irf21.						
DR	GO; GO:0016020; C:membrane; IEA.						
DR	GO; GO:0005634; C:nucleus; IEA.						
DR	GO; GO:0005524; F:ATP binding; IEA.						
DR	GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.						
DR	GO; GO:0005525; F:GTP binding; IEA.						
DR	GO; GO:0007059; P:chromosome segregation; IEA.						
DR	GO; GO:0006810; P:transport; IEA.						
DR	InterPro; IPR003439; ABC transporter.						
DR	InterPro; IPR005289; GTP-binding.						
DR	InterPro; IPR003405; SMC C.						
DR	InterPro; IPR010935; SMC_hinge.						
DR	InterPro; IPR003395; SMC_N.						
DR	Pfam; PF02483; SMC C; 1.						
DR	Pfam; PF06470; SMC hinge; 1.						
DR	Pfam; PF02463; SMC N; 1.						
DR	TIGRFAMs; TIGR00650; MG442; 1.						
KW	Complete proteome.						
SQ	SEQUENCE 1189 AA; 137542 MW; 0P6F9563BE31BC8 CRC64;						

Qy 5 VADKTVEVVK-----NAITAGALDLYNKYLDQVTPWOTFDTTIELKLSRFKQEYSQAA 58
 Db 161 IIBESAGLYKKRKAESIOKLHTEONLNRVEDILDYLGREVPFLKEAAIAKEYKOLS 220

QY 59 S-----VLVDGIKTLMSOD-----KYFATQTVYEWGCVATQALLAAYILLFDEYN 105
Db 221 KEMEQSVDIVTVSDIDHYEDNQDLRLNHLKSQAQKE--GQAQINO-----LLQKYK 274
QY 106 EKKAQAQDILIKVLDGDKTKEAQSLLVSSQSFNNASGKLLALDSQITND-----F 159
Db 275 GKQONDYD-----IEKLN--YELVKATENYEQLSGKLVLEERKQKQSETNARY 322
QY 160 SEKSYFSQSDVKIRKAYAGAAGVVGFGFLIISYSIAAGVVEGKFLIPELKN-----213
Db 323 EEBLDNLESQIDSINKNEK-----AQNEKLLAELKNKQKQLN 358
QY 214 -----KLKSVONFPTLNTVVKQAKDI-----DAAKLKT 244
Db 359 KEVQESLSYISDEQHEKLEEKINSYITLMSEQSDVNDIRPLEHTINENAKSRLD 418
QY 245 TEIAAIGEIKETETTFYVDYDDLMLSLKKEAAKMNCTNCEYQKRHGKTLFVPE 302
Db 419 SLRVE-----AFNLKDIQONITQOKEYQS--SKKSMKEVQ 454
RESULT 29
QY Q9BJY0 PRELIMINARY; PRT; 2752 AA.
AC Q9BJY0;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE 235 kDa rhothry protein (Fragment).
GN Name=IIIA.1;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21273138; PubMed=11378199;
RA Khan S.M., Jarra W., Bayele H., Preiser P.R.;
RT "Distribution and characterisation of the 235 kDa rhothry multigene
RT family within the genomes of virulent and avirulent lines of
RT Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 114:197-208(2001).
DR EMBL; AF233442; AAAK15625.1; .
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON_TER 1
FT NON_TER 2752
SQ SEQUENCE 2752 AA; 322608 MW; E834E3753168AF76 CRC64;
Query Match 8.2%; Score 123; DB 2; Length 2752;
Best Local Similarity 21.3%; Pred. No. 86;
Matches 72; Conservative 58; Mismatches 108; Indels 100; Gaps 16;
QY 3 EIVAD-----KTVE-VVKNAIETADGALDY--NKYLDQVWPOTFDETIKELSRF 50
Db 1140 EKVADINTYXPEKEIEKKIENVVKKKKKKIYKINKLINEISEIEKDKTSLELXNI 1199
QY 51 KQYSAASVVLVDGIKTLMSODKDYFEATQTVYEWGCVATQALLAAYILLFDEYNEKAS 110
Db 1200 NLSYGRS-----LGNIFLEQIDEKKAERT-----IKAMEAYIEDLNIKKKSD 1245
QY 111 AQKDILIKVLDGDKTKEAQSLLVSSQSFNNASGK-----LLALDSQITNDPS 160
Db 1246 IEKDMKIKM-----DINEEMKALNINDDRRNHYTKSKDKHKGSDIDHKSQKIQNPS 1299
QY 161 EKS-----SYFQSDVKIRK-----EAYAGAAGVVGFGFLIISYSIAAGVVEGKFLIP 209
Db 1300 KESDINNKNELQENYSESKHNSDINHLYSKVENI-----YNI-----1338
QY 210 ELK-NKLSQVONFPTLNTVVKQAKDI-----RAKLTTEIAAIGEIKETETTF 259
Db 1339 -LKNLKIIDIYKXEYDEIBKNNKINSINDELNSGKILTKIK---ENSSLTECOSKIES 1394
QY 260 T-----RFVVDYDDLMLSLKKEAAKMNCTNCEYQK 290

Db 1395 TIDNYSIKCIKDIADLTKTYLSEE-----NNINTYLK 1427
RESULT 30
QY Q71X69 PRELIMINARY; PRT; 927 AA.
AC Q71X69;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Membrane protein, putative.
GN OrderedLocusNames=LMOF2365_2330;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=265669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Niernan W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Kathariou S., Wondolung L.D., Uhlich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
RT borne pathogen Listeria monocytogenes reveal new insights into the
RT core genome components of this species.";
RL Nucleic Acids Res. 32:2386-2395(2004).
DR EMBL; AB017330; AAT05096.1; .
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR001809; Outrsurface.
DR Pfam; PF00015; MCPsignal; 1.
DR ProDom; PD001127; Outrsurface; 1.
KW Complete proteome.
SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;
Query Match 8.1%; Score 122.5; DB 2; Length 927;
Best Local Similarity 21.7%; Pred. No. 26;
Matches 75; Conservative 35; Mismatches 137; Indels 99; Gaps 12;
QY 3 EIVADKTVEVVKNAIETADGALDYLNKYLDQVWPOTFDETI-----KELSRFQKQYSQAA 58
Db 166 EAIFDKIKESGEGFAQAADGS-----GKIKDGLVKSQEGNKTISTNLKTLADSSLTFKDGA 221
QY 59 SVL-----VGDIK-----TLLMDSQDKYF 77
Db 222 NTLVEGLKTYTDGVTNTAAAGGDKLNDGVSTLAAGVGLPKDGVAAALDGGATKLASGVSTYT 281
QY 78 EATQTVYEWGCVATQALLAAYILLFDEYNEKKAQAQDKIILIKVLDGDKTKEAQSLLVS 137
Db 282 SGVDTL---AGGINQATYGTSTALSDGLNKNKS-----VPTLASGITQLNNGQKSLATG 332
QY 138 SQSFNNASGKLLA-----LDSQITN-----DFSEKSYFSQSDVKIRKAYAGAAGVVA 187
Db 333 LDSLVDSGNSKLISAGLKELDGNLTDKQGIKQALQAGMNDLQOQIDQLNKSVNGBDAA-----388
QY 188 GPFGLIISYSIAAGVVEGKFLIPKKNLKSQVONFPTLNTVVKQAKDIIDAAKLTTEI 247
Db 389 -----LAKQLAA-----LQKSLDGLQNLGTLFIKS---NANFDAEAKSKINATA 429
QY 248 AAIGEIKETETTFYVDYDDLMLSLKKEAAKMNCTNCEYQKRHG 293
Db 430 GVSADKQKI-----IDAIQADLDKETQKTSATQVATVEQLQSG 467
RESULT 31
QY AAT05096 PRELIMINARY; PRT; 927 AA.
AC AAT05096;
DT 10-MAY-2004 (TRENBLrel. 27, Created)
DT 10-MAY-2004 (TRENBLrel. 27, Last sequence update)

10-MAY-2004 (TrEMBLrel. 27, Last annotation update)
 DE Membrane protein, putative.
 GN LMOF2365_2330.
 OS Bacteria; Monocytogenes str. 4b F2365.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria;
 OC Listeria monocytogenes.
 CC NCBI_TaxID=265669;
 RN [1]
 EN SEQUENCE FROM N.A.
 RC STRAIN=4b F2365;
 RX PubMed=15115801;
 RA Nelson K.B., Fouts D.E., Mongodin E.P., Ravel J., DeBoy R.T.,
 RA Kolonay J.F., Raeko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
 RA Peterson J., White O., Nelson W.C., Niemeyer W., Beanan W.J.,
 RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
 RA Hatt D.H., Selengut J., Van Aken S., Khouri H., Fedorova N.,
 RA Forberger H., Tran B., Kathariou S., Wondolting L.D., Unlich G.A.,
 RA Bayles D.O., Luchansky J.B., Fraser C.M.;
 RA "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
 borne pathogen *Listeria monocytogenes* reveal new insights into the
 core genome components of this species";
 RL Nucleic Acids Res. 32:2386-2395(2004).
 DR EMBL; AE017330; AAT05096.1; -;
 SQ SEQUENCE 927 AA; 97232 MW; 9AE73C67B14685CA CRC64;

Query Match 8.1%; Score 122.5; DB 2; Length 927;
 Best Local Similarity 21.7%; Pred. No. 26;
 Matches 75; Conservative 35; Mismatches 137; Indels 99; Gaps 12;

QY 3 EIVADKTVVVKNAIETADGALDLYNKYLDQVWPQTFDETIK-----KELSRFKQYEQAA 58
 DB 166 EAIFDKIKESGEGFAADGS-----GKIKGLVKSQEGNKITSTNLKTLADSSLTFFKDA 221
 QY 59 SVL-----VGDIK-----TLLMDSQDKYF 77
 DB 222 NTLVGLKTYTDGVTNTAAAGDKLNDGVSTLAAGVPLKQGVAAALDGGATKLSGVSTYT 281
 QY 78 EATQTVVWCQVATQLAAAYILLDFEYNEKASQAKDILIKVLDDGTTKLENAQKSLVS 137
 DB 282 SGVDTL---AGGINQAVTGSTALSDGLNKNMGS-----VPTLASGITQLNNGQKSLATG 332
 QY 138 SQSFNNASGKLLA---LDSLTN-----DFSEKSSYFQSQVDKIRKEAYAGAAAGVVA 187
 DB 333 LDSLVDGSKNLGSLKELDGLNLTQKQIAQLKQGMNDLQGIQDNLKNSVNGEDAA----- 388
 QY 188 GPFGLIISYSTAGVVEGKLIPELKNLKSQNFETTLSTNTVKQAKDIDAAKLKLATTEI 247
 DB 389 -----LAKQLAA-----LQKSLGDLQNGLTPIKS---NANFDAEAKSKINATA 429
 QY 248 AAIQEIETETTRFYVDYDDMLSLKLEAKKRMINTCNEYQKRGH 293
 DB 430 GVSADKQKI-----IDAIQADLDKETOKSATQVATVEQLQSG 467

RESULT 32
 Q7SH24
 ID Q7SH24 PRELIMINARY; PRT; 4007 AA.
 AC Q7SH24
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=NCU00658.1;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 CC NCBI_TaxID=5141;
 RN [1]
 EN SEQUENCE FROM N.A.
 RC STRAIN=OR74A.
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
 RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Reiman B.,
 RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,

RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
 RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
 RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
 RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
 RA Kamal M., Kamvysellis M., Mauceli E., Bielke C., Rudd S., Friseman D.,
 RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
 RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmari S.A.,
 RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
 RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
 RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
 RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
 RT "The Genome Sequence of the Filamentous Fungus *Neurospora crassa*";
 RL Nature 0:0-0(2003).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; AABX01000001; EAA36562.1; -;
 DR InterPro; IPR009638; Fez1.
 DR InterPro; IPR003900; KID repeat.
 DR InterPro; IPR001638; SBP_bac_3.
 DR Pfam; PF06818; Fez1; 1.
 DR Pfam; PF02524; KID; 5.
 DR PROSITE; PS01039; SBP_BACTERIAL_3; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 4007 AA; 453246 MW; 67CDF6E6F1463612 CRC64;

Query Match 8.1%; Score 122.5; DB 2; Length 4007;
 Best Local Similarity 21.4%; Pred. No. 1.4e+02;
 Matches 73; Conservative 57; Mismatches 132; Indels 79; Gaps 13;

QY 11 EVKNAIETADGAL-----DLYNKYLDQVWPQTFDETIK-----ELSRFKQYEQ 56
 DB 1159 EILKQKQSVGSESIATIKYKDKLDELNRNNTSDAIKLKHENELANPKAYEQ 1218
 QY 57 AASLVGDIITLMDSDQKFEATQTVYVCGVATQLAAVILLDFEYNEKKA----- 109
 DB 1219 EKQLAVQHKTEMSLTDRYHEKE-----LATQYQERVQALSAELADKKTALAEYKE 1271
 QY 110 --SAQKDLIKVLDDGTTKLENAQKSL-----LVSSQSFNNAS-----GKLLALD--- 152
 DB 1272 QLSASRAQLDKLADHGVDELQAKLSEVAKVTADYEGNLSLRTKHQGEVNVLVHH 1331
 QY 153 ----SQTNDPSEKSSYFQSQVDKIRKEAYAGAAAGVAGPFGILLIISYSTAGVVEGKL- 207
 DB 1332 QDEIKKLTAGNEKIRNLEHRLINDKAEKQDRA-----BFDKKKALLEGEVA 1379
 QY 208 -----IPELKNLKSQNFETTLSTN---TVQAKNDIDAAKLKLTTTEIAAIGIKTETE 258
 DB 1380 TLOGKVDKSKSLSSKEAEFNLKLENAQIAELRKDVADKSNLSQDKLELSDLKGQOK 1439
 QY 259 T--TRFYVDYDDMLSLK-----EAAKRMINTCNEYQKR 291
 DB 1440 TRIEDFNVQINERMAQLKKAQNELKASQASLNTTTTTEYDAK 1480

RESULT 33
 Q81754
 ID Q81754 PRELIMINARY; PRT; 852 AA.
 AC Q81754
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Rhostry protein (Fragment).
 GN Name=rhop235-8;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 CC NCBI_TaxID=73239;
 RN [1]
 EN SEQUENCE FROM N.A.
 RP Khan S.M., Jarra W., Bayele H., Preiser P.R.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

```

RA Green J.L., Holder A.A.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Baye H.K.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ489163; CAB32953.1; -.
FT NON TER 1
FT NON TER 852
SQ SEQUENCE 852 AA; 99060 MW; 026995B7A6B658B1 CRC64;

Query Match      8.1%; Score 122; DB 2; Length 852;
Best Local Similarity 21.1%; Pred. No. 25;
Matches 71; Conservative 61; Mismatches 108; Indels 96; Gaps 16;

Qy 3 EIVAD-----KIVE-VVNALETADGALDLY---NKYLDQVIPWQTFDETIKLSRF 50
Db 1 EKVDATINYEDPEIEKKIENVVKDKKNYIYKEINKLNEISEIEKDKTSLEELKNI 60
Qy 51 KQEYSQAASVLGDIKTLMDSDQKYFEATQTVYVEMCGVATQLLAAAYLLFDEYNEKAS 110
Db 61 NLSYGRS---LGNIFLEQIDEKKKAERT-----IKAMEAYIEDLDNIKKKSDE 106
Qy 111 AQDILIKVLDDGITKLENAQKSLVSSQSFNNASGKLLALDSQL--TNDFSEKSSYP--- 166
Db 107 IEKDMKIKM-----DINEEMKALNISNDDRNYHTK--SKDHKKGISDIHDKSSKIION 158
Qy 167 --OSQVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEG-KLIPELKNKLSKVONFF 222
Db 159 FSKESDINNITE-----LQENVSESQKNSDINHLSKVENIY 197
Qy 223 TTL-----SNTVKQANKDIDAACL---KLTFEI---AAIGETETETTT- 260
Db 198 NILKLNKIKKIIDVKVETDIEIKNNKSINDELINSKGIITIKENSLSLTCQSKIESSTI 257
Qy 261 -----RFYVDYDILMLSLLEAKKMNINTCNEYQK 290
Db 258 DDNYISKICKIDADLKTYLSE-----NNINTYLR 288

RESULT 34
Q9SZK7 PRELIMINARY; PRT; 1496 AA.
AC Q9SZK7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein F20D10.190 (Hypothetical protein AT4g38070).
GN Name=F20D10.190; Synonymus=AT4g38070;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Bevan M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035538; CAB37547.1; -.

```

```

DR EMBL; AL161592; CAB80472.1; -.
DR FIR; T05634; T05634. HLH_basic.
DR InterPro; IPR001092; HLH_1.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS50888; HLH; 1.
KW Hypothetical protein.
SQ SEQUENCE 1496 AA; 173346 MW; 03BBB75DCD2B03FE CRC64;

Query Match      8.1%; Score 122; DB 2; Length 1496;
Best Local Similarity 18.9%; Pred. No. 49;
Matches 64; Conservative 72; Mismatches 124; Indels 78; Gaps 14;

Qy 4 IVADKTVVVKNAIETADGALDLYNK--YLDQVIPWQTF-----FDETIKLSRF---KQ 52
Db 153 IEAEKTVKGMKGRDGVVVKMEESQVVEEKWKKEQFKLEEAAYEKLKFLKDSKK 212
Qy 53 EYQAASVLGDIKTLMDSDQKYFEATQTVYVEMCGVATQLLAAAYLLFDEYNEKKA 109
Db 213 EWEEESKLLDEIYSIQTKLDSVTRISEDQKQLQMCNGALTO-----BETRKHIL 263
Qy 110 SAQDILIKVLDDGITKLENAQKSLVSSQSFNNASGKLLALDSQLTNDFSEKSSYPQ-- 167
Db 264 EIQVSEFKAKYEDAFACQDARTQL-----DDLAKRDWEVAELRQTLMSKDAYFKEM 316
Qy 168 ----SQVDKIRKEAYAGAAAGVAGPGLIISYSIAAGVVEG---KLIPELKNKLSKVON 220
Db 317 KYENGKLEQENRELLGSLK-----ELQEATIQSGNSALSCLKNKNFRLEN 362
Qy 221 FFTL-----SNTVKQANKDIDAACLKLITTEIAAIGIKTETE-----TTRFYV 264
Db 363 IHKNCANLSKEAEWSSQVKEWVEHNDYKLOQSKAAALKEVELENCRSSTAKMRL 422
Qy 265 DYDDLMLSL-----KEAKKMNINTCNEY---QKRHG 293
Db 423 QYEEISIMFLVLSRTVSEAQSRLANAKDKQIKDEKREG 460

RESULT 35
Q7RFQ7 PRELIMINARY; PRT; 1081 AA.
AC Q7RFQ7;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 235 kDa rhostry protein (fragment).
GN Name=PY04647;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallen S.J., van Aken S.B., Riedmuller S.B., Feidlyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002)
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001425; EAA16541.1; -.
FT NON TER 1
SQ SEQUENCE 1081 AA; 126062 MW; DDFPFE8EFD6D6A94 CRC64;

```

Query Match 8.1%; Score 121.5; DB 2; Length 1081;
 Best Local Similarity 19.1%; Pred. No. 36;
 Matches 74; Conservative 72; Mismatches 129; Indels 113; Gaps 17;

QY 4 IVADKTEVVKNAIETAD-----GALDLYNKYLDOVIPWOTFDTETIKELSRFKQY 54
 DB 589 LVACSTLINENIEAQNDEQIKKHLGELDSKYDKINNI-----TFDEISKVLNNVNNK- 643

QY 55 SQAASVLVGDIKTL-----LMSDQKYPEATQTVYVCGVATQLLAAAYILLFD----- 102
 DB 644 KEALLELINEIKTYVYHMINDLSSILNSVQTDKHK-----LQSSIDSYGYLYNELKNYKEE 700

QY 103 -----BYNEKKA-----SAQDILIKVLDDGKITKLEAKQSKLLYS-----SQSF 141
 DB 701 ILKKNMDYEDYKSKILEYSNNFISKENIILKIDIEE-ISKISNEVKTSLPKYDNEVKKIY 759

QY 142 NNASGKLLALDSOL-----TNDFSEKSSYFOSQVDKIRKEAYAGAAA 183
 DB 760 PNSNGEYLVQVKLIEQIEKYTSBYLKNYKNKNDIKTVSDTKQIQITNKRMDNYKVL 819

QY 184 GVVAGPPLIISYIAAGVVEGK--LPELKNLKSQVNFFTLSNTVQOA-----NKDI 236
 DB 820 NTVMQVQSI--YELIKNIINGKNVLDLRLQSIQAION-YDNINETVKNSSITLNNKI 876

QY 237 DAALKLTTEIAAIGETETETFRY-----VDYDDLMSL----- 273
 DB 877 SMIDDKLNT--TTINDIETKLSTLKEYFTSTKQIQIMENSEIEDNLLQNVNEIDOVQNST 934

QY 274 -----LKEAKKMINTCNEYQKH 292
 DB 935 NALNEQYISLMDVEKLINDTNEIRKH 962

RESULT 36
 Q7RAW8 PRELIMINARY; PRT; 1116 AA.

AC Q7RAW8
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE P235 rhotry protein B5 (Fragment).
 GN Name=PY06381;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNI;
 RX PubMed=12368865;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Perteza M., Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L., Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L., Shallow S.J., van Aken S.E., Riedmuller S.B., Feidblyum T.V., Chao J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M., Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A., Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
 CC EMBL; AAL01002149; EAA18583.1; --
 FT NON_TER 1116 1116
 SQ SEQUENCE 1116 AA; 131905 MW; 3B264B3B5701B222 CRC64;

Query Match 8.0%; Score 121; DB 2; Length 1116;
 Best Local Similarity 21.0%; Pred. No. 40;
 Matches 72; Conservative 57; Mismatches 134; Indels 80; Gaps 17;

QY 7 DKTVEVVKNAIETADGALDLYNKYLDOVIP-WOTFDTETIKELSRFKQYEQSAASVLVGDI 65
 DB 460 ENTLDYFKNK-RNRDINSNDI-NKNILLIPLMDQFDTLANKSMUKLKNDAASEKYVIITQI 517

QY 66 KTLMDSD--QDK---YFEATQTVYVCGVATQLLAAAYILLFDYNEKKAQAQDI--LIK 118
 DB 518 KQKLNOSTYDDKKTGTFSALKAESW---ETKELETIAKLNK-NEETVKLENEIRDLFK 573

QY 119 VLDDGIT-----KLNEAQSKLLSVSQSPNNAAGKLLALDSQLTN----- 157
 DB 574 KYSDEVAEKYIIEELKCLKETIKDIYKKYKIK-----RAIDLKKEIKENKNYIDELGKK 629

QY 158 -----DFEKSYSFOSQVDKIRKEAYAGAAVAGPFGLLIISYIAAGVVEGKLIPEL 211
 DB 630 TPFQIEBYVKKDDTYITKTSELSEYK-----NIIELYNELSSVQENTIDPI 679

QY 212 KKK-----LKS-VONFETLSN---TVQANKDIDAALKLTTEIAAI-----GEIKTE 256
 DB 680 KKKLELTLSKSEIDNVYKQNKIEVELHLKNIETSKNELSNLILEIKIYFGEIDND 739

QY 257 TETFRFYVDYDDLMSLLEAKKMINTCNEYQKHGKTLFE 299
 DB 740 LNKI-----LKEFKNKEQELSKINDYTKENDQLSVYQ 772

RESULT 37
 Q8N592 PRELIMINARY; PRT; 961 AA.

AC Q8N592
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Vesicle docking protein p115.
 GN Name=VDP;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullah S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC032654; AAB32654.1; --
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008565; F:protein transporter activity; IEA.
 DR GO; GO:0006886; P:intracellular protein transport; IEA.

```
DR InterPro: IPR008938; ARM.
DR InterPro: IPR000225; Armadillo.
DR InterPro: IPR006955; Uso1_p115_C.
DR InterPro: IPR006953; Uso1_p115_head.
DR Pfam: PF04871; Uso1_p115_C; 1.
DR Pfam: PF04869; Uso1_p115_head; 1.
DR PROSITE: PS50176; ARM_REPEAT; 1.
DR SEQUENCE 961 AA; 107765 MW; DCF4888F4B8C02A CRC64;

Query Match      8.0%; Score 120.5; DB 2; Length 961;
Best Local Similarity 23.7%; Pred. No. 36;
Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTVEVVKNAIETADGALDLY-NKYLDQVTPWOTFDETIKELGRFKQEVQAASVLVGI 65
Db 634 DKKEEVVKTLQEHNDIVTHYKNMIREQDLQLELRQOVSTLKQNEQLQTAQTAVTQVSOI 693
QY 66 K-----TLIM-----DSQDKYFEATQVYVCGVATQLLAAVILLFDEYNEKKAS-- 110
Db 694 QQHKDQYNLLKIQLGKDNQHGYSSEGAQ-----MNGIQPEEIGR---LREEIELKRNQE 746
QY 111 -----AQKDILIKVLDDGITKLENAQKSLVS---SQSFNNASGKLLALDLSQLTNDFSE 161
Db 747 LLOSQLTEKDSMIENKSSQTSCTNEQSSAIVSARDSEQVAELKQELATLKSQI-NSQSV 805
QY 162 KSSYFQSQVDKIRKEAYAGAAVAGVAGPFGGLIISYIAAGVVEGKL-----IPELKNKL 215
Db 806 EITKLQTEKQELLQKTEAFKSVQVQETETIIATKTTD--VEGRLSALLQETKELKNEI 863
QY 216 KSVQNFPTLNTVKQANKDIDAAKLKLTTETIAAIGIKETETTRFYVDYDDMLSLK 275
Db 864 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDSKKEQDQLLV-LLA 913
QY 276 EAAKQKMTNCE 287
Db 914 DQDQKILSLKNK 925

RESULT 38
VDP_HUMAN
ID VDP_HUMAN STANDARD; PRT; 962 AA.
AC O60763;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE General vesicular transport factor p115 (Transcytosis associated protein) (TAP) (Vesicle docking protein).
DE Name=VDP;
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITE SER-942.
RX MEDLINE=98148093; PubMed=9478999;
RA Solda M., Misumi Y., Yano A., Takami N., Ikehara Y.;
RT "Phosphorylation of the vesicle docking protein p115 regulates its association with the Golgi membrane."
RL J. Biol. Chem. 273:5385-5388(1998).
CC -!- FUNCTION: General vesicular transport factor required for intercellular transport in the Golgi stack; it is required for transcytotic fusion and/or subsequent binding of the vesicles to the target membrane. May well act as a vesicular anchor by interacting with the target membrane and holding the vesicular and target membranes in proximity (By similarity).
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein which recycles between the cytosol and the Golgi apparatus during interphase.
CC -!- DOMAIN: Composed of a globular head, an elongated tail (coiled-coil) and a highly acidic C-terminal domain.
CC -!- PTM: Phosphorylated in a cell cycle-specific manner; phosphorylated in interphase but not in mitotic cells.
CC Dephosphorylated protein associates with the Golgi membrane; phosphorylation promotes dissociation.
```

```
CC -!- SIMILARITY: Belongs to the VDP/Uso1/YBL047C family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D86326; BAA25300.1; -.
DR MIM: G03344; -.
DR GO: GO:000139; C:Golgi membrane; TAS.
DR InterPro: IPR008938; ARM.
DR InterPro: IPR000225; Armadillo.
DR InterPro: IPR006955; Uso1_p115_C.
DR InterPro: IPR006953; Uso1_p115_head.
DR Pfam: PF04871; Uso1_p115_C; 1.
DR Pfam: PF04869; Uso1_p115_head; 1.
DR PROSITE: PS50176; ARM_REPEAT; UNKNOWN 1.
DR Coiled coil; Golgi stack; Membrane; Phosphorylation;
KW Protein transport; Transport.
FT DOMAIN 1 637 Globular head.
FT DOMAIN 638 930 Coiled coil (Potential).
FT DOMAIN 935 962 Asp/Glu-rich (acidic).
FT MOD_RES 942 942 Phosphoserine.
FT MUTAGEN 942 942 S->A: Loss of phosphorylation.
SQ SEQUENCE 962 AA; 107906 MW; 2E748F2C1BC2B942 CRC64;

Query Match      8.0%; Score 120.5; DB 1; Length 962;
Best Local Similarity 23.7%; Pred. No. 36;
Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;

QY 7 DKTVEVVKNAIETADGALDLY-NKYLDQVTPWOTFDETIKELGRFKQEVQAASVLVGI 65
Db 635 DKKEEVVKTLQEHNDIVTHYKNMIREQDLQLELRQOVSTLKQNEQLQTAQTAVTQVSOI 694
QY 66 K-----TLIM-----DSQDKYFEATQVYVCGVATQLLAAVILLFDEYNEKKAS-- 110
Db 695 QQHKDQYNLLKIQLGKDNQHGYSSEGAQ-----MNGIQPEEIGR---LREEIELKRNQE 747
QY 111 -----AQKDILIKVLDDGITKLENAQKSLVS---SQSFNNASGKLLALDLSQLTNDFSE 161
Db 748 LLOSQLTEKDSMIENKSSQTSCTNEQSSAIVSARDSEQVAELKQELATLKSQI-NSQSV 806
QY 162 KSSYFQSQVDKIRKEAYAGAAVAGVAGPFGGLIISYIAAGVVEGKL-----IPELKNKL 215
Db 807 EITKLQTEKQELLQKTEAFKSVQVQETETIIATKTTD--VEGRLSALLQETKELKNEI 864
QY 216 KSVQNFPTLNTVKQANKDIDAAKLKLTTETIAAIGIKETETTRFYVDYDDMLSLK 275
Db 865 KALSEERTAIKEQLDSSNSTIAI-----LQTE-----KDKLEITDSKKEQDQLLV-LLA 914
QY 276 EAAKQKMTNCE 287
Db 915 DQDQKILSLKNK 926

RESULT 39
Q7RD43
ID Q7RD43 PRELIMINARY; PRT; 1611 AA.
AC Q7RD43;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
DE Name=PY05582;
GN Plasmodium yoelii yoelii.
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
```

```
RX PubMed=12368865;
RA Carlton J.C., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.C., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdey A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01001787; EAA17620.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1611 AA; 188293 MW; CD0493CC4A9BC5D1 CRC64;

Query Match      8.0%; Score 120.5; DB 2; Length 1611;
Best Local Similarity 22.3%; Pred. No. 66;
Matches 75; Conservative 60; Mismatches 120; Indels 81; Gaps 17;

QY 8 KTVVVVNAIETADGA-LDYNKYLDQVWPQTFDETIK-ELSRFQOEY-----SQASVL 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
784 KNKLKK--MSTYKGYCNIRNLNL-----MKNEKIKFENKLLKDYDNLSKYKNVL 835
QY 62 VGDIKTLMM-----DSQKYFEATQTVYWCVGATOLLAAYILLFDEYNEKASQAQ 112
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
836 VLCKKIFLKYTTKDRSYNDKYE-----NNMRHLQSEKNLLNIIEEKYQE 886
QY 113 KDILIK-VLDDGITKLNKAEQSLVSSQSNFNASGKLLALDSQLTNDPFS-KSYFQSQV 170
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
887 NNININMLSDELKQLQENESIISNNNSKYTEVITNSKYQHLQDNFKIKSEHEKLI 946
QY 171 D-----KIRKAYAGAAAGVVGPFGLIISYIAAGVVEG-----KLIPELKN 213
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
947 EHNKIKRENENIKETLIKELGDTKAKYFNM-----TGILQBEKMYAKKIKEL 999
QY 214 KLASVQNFPTLNTYKQAKDIDA-AKL-----KLTTEIAIGEIKTETETRFY 263
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1000 KLOKEBEKNNIINDVKD-----EIDFTFKILEKENENHKKELKELNIEEKYEQIN 1056
QY 264 VD-----YDDLMLSLKLEAAKGMINTCNEYQK 290
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1057 FDSLKKEFEKSFDEVQI-ILKEMIQKEKYTNSYNK 1091

RESULT 40
Q6GNE7 PRELIMINARY; PRT; 1012 AA.
ID Q6GNE7
AC Q6GNE7
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schenman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.B., Scheetz T.E.,
```

```
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S.G., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: BG073565; AAH73565.1; -.
DR InterPro: IPR009053; Prefoldin.
DR InterPro: IPR007794; Rib_recept_KP.
DR Pfam: PF05104; Rib_recpt_Rp_reg; 1.
KW Hypothetical protein.
SQ SEQUENCE 1012 AA; 112965 MW; CF23AB8E0D9EDAEC CRC64;

Query Match      8.0%; Score 120; DB 2; Length 1012;
Best Local Similarity 21.5%; Pred. No. 41;
Matches 70; Conservative 58; Mismatches 124; Indels 74; Gaps 14;

QY 12 VVKNAIETADGALDLYNKYLD-QVWPQ-----TPDETIKELSRFQOEYSQAASV 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
669 IQKEAAELKAEVQKKNLDRKNQWAMEALGLAEKCEKLNSEKKAKEEMVQQLSA 728
QY 61 LVGDIKTLMD-----SDQKYFEATQTVYWCVGATOLLAAYILLFDEYNEKASQAQ 112
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
729 VOSQTKETLQSVLPQITIVSQSYSEWLQ---BFRERTSQLLS-----QOTEKEGSSE 778
QY 113 KDILIKVLDGDIKLNKAEQSLVSSQSNFNASGKLLALDSQLTNDPFSK-----SSYFQS 168
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
779 LQLQLKQAE-----SQSDLVCEKRYTILQTEAMKALQNSVEESEQWKAFFSS 831
QY 169 QVDKIRKAYAGAAAGVVGPFGLIISYIAAGVVEGKLIPELKNKLSVQNFPTTILSNT 228
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
832 SEELK-----SHSQVKLEE--TVEKLSRDLQSTEQLEKCVSLM 870
QY 229 VKQAKNDIDAAKULKT---EIAAIGEIKTET-----ETTFYVDYDDLMLSLKKAQKM 281
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
871 EAQLETQMAKSTECQTSNEIESLQALLSEHLDATKAEARKOSIELSVLRQQLGEM 930
QY 282 INTCEYQKEHG---KKTLE---FEVPE 302
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
931 LNHVNDTEK-HGTEEHQTLQNKEVPK 955

RESULT 41
MYSC_CHICK
ID MYSC_CHICK STANDARD; PRT; 1102 AA.
AC P29616;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Myosin heavy chain, cardiac muscle isoform (Fragment).
OS Gallus gallus (Chicken).
```



```
Db 1363 AKTELSDLITSSNQAPADAIS-----DAQKTLDEI 1393

RESULT 43
CAF32691 PRELIMINARY; PRT; 1828 AA.
AC CAF32691
DT 31-MAR-2004 (TrEMBLrel. 27, Created)
DT 31-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 31-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Variable membrane protein precursor.
GN VMP.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4195;
RA Boesen T.;
RT "Gene and protein structure of the Mycoplasma hominis vaa adhesin.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ629113; CAF32691.1;
KW SIGNAL.
FT CHAIN 1 28 Potential.
FT CHAIN 29 1828 variable membrane protein.
SQ SEQUENCE 1828 AA; 205343 MW; 6A0D98F6491A1570 CRC64;

Query Match 8.0%; Score 120; DB 2; Length 1828;
Best Local Similarity 19.8%; Pred. No. 82;
Matches 79; Conservative 53; Mismatches 111; Indels 156; Gaps 14;

QY 2 TEIVADTKVEVKNVNAITAGAL-----DLVNYKLD-----QVTPWQTFD 41
Db 1051 TKALEDKSELAQNIIDTKNNATKTFNDKDELNKLIDTSDAKSVNKKESDVLKONSID 1110
QY 42 ET--IKELSPKQEYSAASVILGDIKTLMDSDQKYFEATQTYEHCQVATQLLAAYIL 99
Db 1111 SNTPIKEIKATEKITEAINSLTSIK-----NKKDEE----- 1143
QY 100 LFDEYNKKSAQKDIILIKVLDDGITKLNEAQSLLV-----SSQSFNAS 145
Db 1144 -FNKYNDIKTSLEN--LIKEEDAVOVGIADVQKTLSENNVDKATIEKIQHSTEALTHAK 1200
QY 146 GKLLAL-----DSQLTDFSEKSSYFQ-----SQVDKIRKEAYAGAAAGVAGPPGLI 193
Db 1201 EELKKLIDTTKKQLTKEFETKSELEKLSLIPANNVDDKKDELSIFG----- 1247
QY 194 ISYSIAGVVEGKLIPELKNKLSKVONFTTSLNTV-----KQANKD 235
Db 1248 -----NTNITNSDSIKQIKETIKIQALESNTNISKQKEBLEKYNVTTKTALEQLIKD 1302
QY 236 IDAAKLLTTEIAAI-----GEIKTETET-----TR 261
Db 1303 EDAKEVGTDDANAAITKNKADKNSTLEITNATKALEDKSKLDQELIKTKKEAFNNLTK 1362
QY 262 FYVDYDDLMLSLKKEAAKMMINTCNEYQKRGKKTLEFV 300
Db 1363 AKTELSDLITSSNQAPADAIS-----DAQKTLDEI 1393

RESULT 44
Q6CXQ4 PRELIMINARY; PRT; 655 AA.
AC Q6CXQ4
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to sp|P34237 Saccharomyces cerevisiae YKL179c.
GN ORFNames=KLLA0A06402g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
```

```
RN SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marcq C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boismare A., Boyer J., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Despons L., Fabre E., Fairhead C., Jaumiaux N., Joyet P., Kachouri R.,
RA Hantraye F., Hennequin C., Jaumiaux N., Lesur I., Ma L., Muller H.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382121; CAH02873.1;
SQ SEQUENCE 655 AA; 74898 MW; 8024192DF56618EC CRC64;

Query Match 7.9%; Score 119.5; DB 2; Length 655;
Best Local Similarity 21.3%; Pred. No. 27;
Matches 69; Conservative 56; Mismatches 106; Indels 93; Gaps 13;

QY 24 LDLYNKYLQVTPWQTFDETIKELSR-----FKQEYSAASVL---VGDIKTLMDSDQ 74
Db 91 IELYELKILQPDPTGILESIVEELKDGTVNDDIKAENQKLAIVSKCADIYTLKKRLSD 150
QY 75 KYFEATQTV-----YEWCGVATQLLAAYILLFDEYN---EKKASQK 113
Db 151 LEQNSAKTILNRLIAKEKITSKWEKQKRNWNRRELLKQLETNNTSILEKKIGAQV 210
QY 114 DILIKVLDDG-----ITKLNEAQSLLVSSQSFNASGKLALDS-----QLTN 157
Db 211 D--LEQEDGDDDTVLVSGKNSAQSELLV---QEEAAQIRIMSLESRNEELNSEVTKLK 266
QY 158 DFSEKSSYFQSQVDKIRKEAYAGAAAGVAGPPGLIISYSIAGVVEGKLIPELKNKLS 217
Db 267 D-EQDSIFQEKETKLNQLESENA-----KLVTCLEERKS 301
QY 218 VQNFPTTSLNTVQKA-----NKDIDAACKLKT-----EIAAIGEIK--TETET 259
Db 302 LKETSQTLSQLSSALETQTYKSELETURRKLNTSYSDYEKIQEELNAMKKIEFGTSDDS 361
QY 260 TRFYVDYDDLMLSLKKEAAKMMIN 283
Db 362 DNENDGDDSDITSLLKHANQKLQN 385

RESULT 45
Q86TB8 PRELIMINARY; PRT; 973 AA.
AC Q86TB8
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZp451D234.
GN Name=DKFZp451D234;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
```

RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Oeanger A.,
 RA Fobo G., Han M., Wiemann S.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL832010; CAB89917.1; -.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008565; P:protein transporter activity; IEA.
 DR GO; GO:0006886; P:intracellular protein transport; IEA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR002025; Armadillo.
 DR InterPro; IPR006955; Uo01_p115_C.
 DR InterPro; IPR006953; Uo01_p115_head.
 DR Pfam; PF04871; Uo01_p115_C_1.
 DR Pfam; PF04869; Uo01_p115_head; 1.
 DR PROSITE; PS0176; ARM_REPEAT; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 973 AA; 109186 MW; 9DD0279FC2CC6BFD CRC64;
 Query Match 7.9%; Score 119.5; DB 2; Length 973;
 Best Local Similarity 23.7%; Pred. No. 42;
 Matches 74; Conservative 51; Mismatches 136; Indels 51; Gaps 13;
 QY 7 DKTEVVVKNATETADGALDLY-NKYLDQVTPWQTFDETIKELSRFKQYEQASVVLVGD 65
 DB 646 DKKEEVKVLQEHDDIVTHYKNMIREQDQLBELRQOVSTLKCNEQLQTAVTQVQS 705
 QY 66 K-----TLIM-----DSQDKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKA 110
 DB 706 QQHKDQYNLLKIQLGKDNQHGVSQAQ-----MNGIQPEEIGR---LREEIEELKRNQE 758
 QY 111 -----AQKDLIKVLDGDTKNEAQKSLVS-----SQSFNNSGKLLALDSQLTNDPSE 161
 DB 759 LLQSOLTEKDSMIENKSSQTSQTNQSSAIVSARDSEQVAELKQELATLKSQV-NSQSV 817
 QY 162 KSSVFQSOVDKIRKEAVAGAAAGVAGPFGGLIISYSIAAGVVEGL-----IPELKNKL 215
 DB 818 EITKLQTEKQELLOKTEAFKSVQVQETETIATKTTD--VEGRLSALLQETKELKNEI 875
 QY 216 KSVQNFPTTILSNVQKANKDIDAAKLKLTTEIAAIGBKTEETETTRFYVDYDDMLSLK 275
 DB 876 KALSEERTAIKEQLDSSNTIAI-----LQTE-----KDKLELEITDSKQEDLLV-LLA 925
 QY 276 EAAKMTNCTNE 287
 DB 926 DQDQKILSLKNK 937
 RESULT 46
 Q6BUQ9 PRELIMINARY; PRT; 2042 AA.
 AC Q6BUQ9
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Similar to sp|P25386 Saccharomyces cerevisiae YDL058w US01.
 GN ORFNames=DEHA009658g;
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
 OX NCBI_TaxID=4959;
 RN [1]
 SQ SEQUENCE FROM N.A.
 RC STRAIN=CHS767;
 RG GENOLEVURES;
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
 RA Dufon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
 RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
 RA Boissrame A., Boyer J., Cattolico L., Confanieri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,

RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swenene D., Tekata F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Wincker P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHS767;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CR382135; CAG86127.1; -.
 SQ SEQUENCE 2042 AA; 233328 MW; 959292DEB8EE1158 CRC64;
 Query Match 7.9%; Score 119.5; DB 2; Length 2042;
 Best Local Similarity 19.9%; Pred. No. 1e+02;
 Matches 59; Conservative 60; Mismatches 124; Indels 53; Gaps 10;
 QY 8 KTVVKNATETADGALDLY-NKYLDQVTPWQTFDETIKELSRFKQYEQASVVLVGD 67
 DB 1695 KSISQELSALKSSDKAASEMTKQLENEL--QTLKDDIEEKSRKSELEEKSTLSS 1752
 QY 68 L--LMDSDQKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAQKDLILVLDG 125
 DB 1753 LENKLDAMKELSEKSVIEKLSAELKEHSLKSLADLKEYKEFEQLEKE-----HEQLK 1806
 QY 126 KLNEAQKSLVSSQSFNNSGKLLALDSQLTNDPSEKSSYFQSOVDKIRKEAVAGAA 185
 DB 1807 KKFDAEGR--IHGKMKELKSLDLSQDDLTAAMDLEK-----SKIESLNQEL----- 1851
 QY 186 VAGPFGGLIISYSIAAGVVEGLKIPELKNKLSVQNFPTTILSNVQKANKDIDAAKL 245
 DB 1852 -----LSTKTTKDDIEKLTQLESTQ-----ALKNNKELKDLNLSKENIS 1895
 QY 246 ---EIAAIGBKTEETETTRFYVDYDDMLSLKKE-AAKMTNCTNEYQKRHGK 297
 DB 1896 LKEDLNALKENETRLK-----QDLSEKANEKSLKSSQKNEKDN 1943
 RESULT 47
 Q9LIW7 PRELIMINARY; PRT; 895 AA.
 AC Q9LIW7
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to an Arabidopsis thaliana chromosome BAC genomic
 DE sequence.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP001111; BAA90502.1; -.
 DR Gramene; Q9LIW7; -.
 SQ SEQUENCE 895 AA; 100520 MW; BAA47BFD70FFAA3 CRC64;
 Query Match 7.9%; Score 119; DB 2; Length 895;
 Best Local Similarity 21.1%; Pred. No. 41;
 Matches 74; Conservative 74; Mismatches 123; Indels 80; Gaps 18;
 QY 1 MTEIVADKTVVKNATETADGALD-----LYNKYLDQVTPWQTFDETIK-----EL 47
 DB 268 IAEVNEK-VEILSSVVRKGLDSTAESSEKNEETELV--KNLESEVSVLKGKLEE 324
 QY 48 SRFKQYEQASVVLVGDITLLMDSDQKYFEATQTVVW---CGVATQLLAAYILLPDEY 104
 DB 325 ARTIEERLAETKLEIKELSEVADAKKAESARQLFEWKHKAGLLEMELEA-VTLSDKF 383

```

Qy 105 -NEKASQAKDILIKVLDDGITKLNKAEQKSLVSSQSFNNASGKLLALD---SQTNDPFS 160
Db 384 KGESLASTTBE-----LGKIQSALQDRESIEYVLKGTALTEVARLLADV 431
Qy 161 EKSYF-----QSOVDKIRKEAYAGAAAGVAGPGLIISYSIAAGV-VEG----- 205
Db 432 ESNEQPDASQOEVLGQTTIDVLNKLKLEAAEAASEA-----LNNEKAANVIEGLTEE 485
Qy 206 --KLIPEL-----KNKLSQVNFPTLNTVQKANKOIDAAKLKLTTE-----IAAI 250
Db 486 NVKLISELNETRDREKEKRAVEDLTAALS---EESDKAEAEHVELSKEDDHEHALAQI 542
Qy 251 GEIKTETETTR-----FVYDDDLMLSLKKAANKMINTCNEYQKR-HGKKT 296
Db 543 GDLKALKSTKSYEVMLDEANDYDITCLRKNVDKLEAVNKNYRECESKET 593

RESULT 48
Q7RELO PRELIMINARY; PRT; 2664 AA.
AC Q7RELO
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Rioptry protein (Fragment).
GN Name=PY05054;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Perle M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.J., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoalbi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001582; EAA17018.1; -.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRPFAMS; TIGR01612; 235kDa-fam; 1.
FT NON TER 2664
SQ SEQUENCE 2664 AA; 312670 MW; FDE72DB05743F4AB CRC64;

Query Match 7.9%; Score 119; DB 2; Length 2664;
Best Local Similarity 22.1%; Pred. No. 1.5e+02;
Matches 90; Conservative 50; Mismatches 122; Indels 146; Gaps 21;

Qy 1 MTEIVADKTVVKNNAETADGALDLYNKYLDQVTPWQTFDETIKE----- 46
Db 819 MTKI-STKBEILKNINEVK-----HMKDNFLDKVNIYTFDDIYKENVDSHDQFTBLAK 873
Qy 47 -----LSRFKQEYSQAASVLVGDITKLLMDSQDKYEATQV-----VEMCGVAT 91
Db 874 KIKTDVASKFNEYEQ-----MLGSKALITQTNKSIIEEYQNTLKKANEYQLCDNTT 929
Qy 92 QLLAAYITLLDFEYNE-----KKASAQKDI-----LIKVLDDGI---TKLNE 129
Db 930 ELIKNFINKNELSEILKNKIESINNTSIEKSYAENFINTLNKTKELDDIFKDAASLNE 989
Qy 130 AQK-----SLLVSSQSFNNASGKLLALDLSQLTNDPFSKSYFOSQV 170

```

```

Db 990 HKTEELKTYFNDLKENLGKSNSTLSQOFN---CK-----EKLFDNIQKNDINKNI 1041
Qy 171 DKIRKEAYAGAAAGVAGPGLIISYSIAAGV---VEGKL-----IPE 210
Db 1042 SNIEKEIYSSI-----HNINEDIENETEKNIKSLNTQVFEKVTKNVANLNE 1087
Qy 211 LKNKLSQVNF-----FTTSLNTVKQ---ANKDIDAAKKLKLTTEI-----AAIG 251
Db 1088 IREKLQK-YNFDDFEKEKNKYTDENWKIKDDVTTLNKKDKNITKL-TIEKNSESYS 1145
Qy 252 EIKTETETTRFYVDYDDLMLSLKKAANKMINTCNEYQKRHGKKTTFE 299
Db 1146 EIKTQDKSEKTYD-KTYIKEEPKEIEKKIKNIITKIDR---KKNIFE 1189

RESULT 49
Q7P7E4 PRELIMINARY; PRT; 335 AA.
AC Q7P7E4
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12).
GN Name=FNW1630;
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=209882;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49256;
RA Karpatrial V., Ivanova N., Anderson I., Resnik G., Bhattacharyya A.,
RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
RA Haselkorn R., Overbeek R., Kyrpides N.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the glyceraldehyde-3-phosphate
CC dehydrogenase family.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABF01000017; EAA24760.1; -.
DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006006; P:glucose metabolism; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR006424; GAPDH-I.
DR InterPro; IPR011596; GAPDH like.
DR InterPro; IPR000173; GAP_dhydrogenase.
DR Pfam; PF02800; Gp_dh_C; 1.
DR Pfam; PF00044; Gp_dh_N; 1.
DR PRINTS; PR00078; G3PDHGRNASE.
DR ProDom; PD007761; GAPDH like; 1.
DR TIGRPFAMS; TIGR01534; GAPDH-I; 1.
DR PROSITE; PS00071; GAPDH; 1.
KW Oxidoreductase.
SQ SEQUENCE 335 AA; 35833 MW; 5089AB90E7D16B61 CRC64;

Query Match 7.9%; Score 118.5; DB 2; Length 335;
Best Local Similarity 22.3%; Pred. No. 14;
Matches 73; Conservative 42; Mismatches 110; Indels 103; Gaps 16;

Qy 7 DKTVEVVKNAIETADGALDLYNKYLDQVTPWQTFD-ETIKELSRF--KQYSQA----- 57
Db 55 DGTIEVTNGFVVGSDISKVFAPKANPEELPWGLDGVLECTGFTGFTSKKABAHKAGA 114
Qy 58 -----ASLVGDIKTLMLMSQDKYFEATQVYWCWGAQTQLLAAYITLLDFEYNEKASQA 112
Db 115 KKVVISAPATGDKLTIVYVNVNDVLTGTETVISGASCTTNC LAP----- 158
Qy 113 KDILIKVLDD--GITKLNKAEQKSLVSSQSFNNASGKLLALDLSQLTNDPFSKSYFOSQV 170
Db 159 ---MAVLNNDKFGII-----EGLMTTIHAYTN-----DQNTLDAPHKKG----- 194

```

